

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: RITA MITRA Examiner #: 77775 Date: 11/23/01
 Art Unit: 1653 Phone Number: 301-605-1211 Serial Number: 09/617099
 Mail Box and Bldg/Room Location: 9B01, 2MT Results Format Preferred (circle): PAPER DISK E-MAIL
4203

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Protein RIM2

Inventors (please provide full names): Susumu Seino, Tadao Shibasaki, 1
Nobuaki Ozaki

Earliest Priority Filing Date: 10/3/1999

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

I request a literature search and a patent search on "Protein RIM2".
 - please see attached informations on technical field of the invention.
 - Claims 3-12 have been elected and will be examined.

Keywords:

Point of Contact:
 Mary Hale
 Technical Info. Specialist
 CMM 12D16 Tel: 303-4258

- RIM Protein and isoform
- interacts with Rab3 (protein)
- regulator of Rab3-dependent synaptic vesicle fusion.
- interacts with GTP/GTP exchange factor (GEF)
- regulation of exocytosis of neurons and endocrine cells
- recombinant vector
- DNA encoding RIM protein.

STAFF USE ONLY

| | Type of Search | Vendors and cost where applicable |
|------------------------------------|---|-----------------------------------|
| Searcher: _____ | NA Sequence (#) _____ | STN _____ |
| Searcher Phone #: _____ | AA Sequence (#) _____ | Dialog _____ |
| Searcher Location: _____ | Structure (#) _____ | Questel/Orbit _____ |
| Date Searcher Picked Up: _____ | Bibliographic <input checked="" type="checkbox"/> | Dr. Link _____ |
| Date Completed: <u>11/30/01</u> | Litigation _____ | Lexis/Nexis _____ |
| Searcher Prep & Review Time: _____ | Fulltext _____ | Sequence Systems _____ |
| Clerical Prep Time: _____ | Patent Family _____ | WWW/Internet _____ |
| Online Time: _____ | Other _____ | Other (specify) _____ |

THIS PAGE BLANK (USPTO)

```

OM nucleic - nucleic search, using sw model
November 21, 2001, 18:56:16 ; Search time 358.58 Seconds
                                (without alignments)
Run on: 8720.369 Million cell updates/sec

```

```
Searched: /500... 1980 D
Total number of hits satisfying chosen parameters:
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
post-processing: Minimum Match 0%
                  Maximum Match 100%
Listing first 45 summaries
```

SUMMARIES

[illegible]

Oligonucleotide D1
Oligonucleotide D1
Oligonucleotide D1
Oligonucleotide D2
Oligonucleotide D2
Oligonucleotide D1
Oligonucleotide D1
Human Doc2-beta gene encoding a su
Gene encoding a su
Sequence encoding
Human protein Ytna
DNA encoding a pro
Human low adenosi
Human low adenosi
Human low adenosi
Human low adenosi
Human low adenosi
Human low adenosi
Murine LOBO hemolo
An ECORI fragment
Kaposi's sarcoma -a
KSHV LTR DNA (nuc)
KSHV long unique C
poly(ADP-ribose) traic
Mouse Doc2alpha ge
Plasmodium falcipar
Exons E', C, B and
Exons D', C, B and
Spinocerebellar at
Mouse inscitol pol
DNA sequence of hu
Nucleotide sequenc
Nucleotide sequenc
Nucleotide sequenc
Nucleotide sequenc
Nucleotide sequenc
Glutamine rich reg

THIS PAGE BLANK (USPTO)

WO200058356-A1.

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

3M nucleic - nucleic search, using sw model

Run on: November 21, 2001, 18:56:16 ; Search time 358.58 Seconds
(without alignments)
8720.369 Million cell updates/sec

Title: US-09-617-099b-2

Perfect score: 4980

Sequence: 1 gctccctaggtgtgttcg.....acatgtgcctcgaagaag 4980

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_0601:*

1: /cgnl_9/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /cgnl_9/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /cgnl_9/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /cgnl_9/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /cgnl_9/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /cgnl_9/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /cgnl_9/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /cgnl_9/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /cgnl_9/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /cgnl_9/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /cgnl_9/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /cgnl_9/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /cgnl_9/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /cgnl_9/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /cgnl_9/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /cgnl_9/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /cgnl_9/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /cgnl_9/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /cgnl_9/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /cgnl_9/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /cgnl_9/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /cgnl_9/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1 | 844.8 | 17.0 | 1302 | 19 | AAV40485 | Human secreted pro |
| 2 | 377 | 7.6 | 2276 | 21 | AAFC6011 | Human secreted pro |
| 3 | 207.6 | 4.2 | 355 | 20 | AAV90372 | EST clone DM118. |
| 4 | 207.6 | 4.2 | 355 | 20 | AAV9630 | EST clone CO618. |
| 5 | 195.6 | 3.9 | 378 | 21 | AACT4934 | Human OREF ORF489 |
| 6 | 93.4 | 1.9 | 936 | 22 | AAFS8252 | Oligonucleotide D1 |
| 7 | 93.4 | 1.9 | 936 | 22 | AAFS8254 | Oligonucleotide D1 |
| 8 | 93.4 | 1.9 | 936 | 22 | AAFS8257 | Oligonucleotide D1 |
| 9 | 93.4 | 1.9 | 936 | 22 | AAFS8259 | Oligonucleotide D2 |
| 10 | 93.4 | 1.9 | 936 | 22 | AAFS8262 | Oligonucleotide D2 |
| 11 | 93.4 | 1.9 | 938 | 22 | AAFS8255 | Oligonucleotide D1 |

| | | | | | | |
|------|------|-----|--------|----|-----------|----------------------|
| C 12 | 92.8 | 1.9 | 936 | 22 | AAFS8252 | Oligonucleotide D1 |
| C 13 | 92.8 | 1.9 | 936 | 22 | AAFS8254 | Oligonucleotide D1 |
| C 14 | 92.8 | 1.9 | 936 | 22 | AAFS8257 | Oligonucleotide D1 |
| C 15 | 92.8 | 1.9 | 936 | 22 | AAFS8259 | Oligonucleotide D2 |
| C 16 | 92.8 | 1.9 | 936 | 22 | AAFS8262 | Oligonucleotide D2 |
| C 17 | 92.8 | 1.9 | 936 | 22 | AAFS8255 | Oligonucleotide D1 |
| C 18 | 55.4 | 1.1 | 2043 | 18 | AAAT79627 | Human Doc2-beta gene |
| C 19 | 50 | 1.0 | 10732 | 21 | AAAI0594 | Gene encoding a su |
| C 20 | 47.6 | 1.0 | 2182 | 9 | AAAN80338 | Sequence encoding |
| C 21 | 47.6 | 1.0 | 2244 | 21 | AAZ44431 | Human protein kina |
| C 22 | 47.6 | 1.0 | 2245 | 16 | AAQ79892 | DNA encoding a pro |
| C 23 | 47.6 | 1.0 | 2245 | 21 | AAAF21404 | Human low adenosin |
| C 24 | 47.6 | 1.0 | 2245 | 21 | AAAF35282 | Human adenosine re |
| C 25 | 47.6 | 1.0 | 38644 | 21 | AAAF21424 | Human low adenosin |
| C 26 | 47.6 | 1.0 | 38644 | 21 | AAAF35302 | Human adenosine re |
| C 27 | 46.6 | 0.9 | 49999 | 20 | AAZ23895 | Murine LOBO homolo |
| C 28 | 45.4 | 0.9 | 1159 | 21 | AAAS9240 | An EcorI fragment |
| C 29 | 45 | 0.9 | 3489 | 21 | AAAS9240 | Kaposi's sarcoma-a |
| C 30 | 45 | 0.9 | 32207 | 20 | AAV73805 | KSHV LUR DNA (nuc |
| C 31 | 45 | 0.9 | 137507 | 19 | AAV19941 | Polyglutamine tirc |
| C 32 | 44.4 | 0.9 | 486 | 22 | AAFT5507 | Mouse Doc2alpha ge |
| C 33 | 44.4 | 0.9 | 2255 | 20 | AAZ00321 | Plasmodium falcipa |
| C 34 | 44.4 | 0.9 | 3579 | 21 | AAAT0099 | Exons E, C and A o |
| C 35 | 43.8 | 0.9 | 1037 | 21 | AAAS9242 | Exons D, C, B and |
| C 36 | 43.8 | 0.9 | 1472 | 21 | AAAS9241 | Spinocherebellar at |
| C 37 | 43.4 | 0.9 | 397 | 20 | AAAB9891 | Mouse inositol pol |
| C 38 | 43.2 | 0.9 | 1876 | 17 | AAAT29743 | DNA sequence of hu |
| C 39 | 43 | 0.9 | 5120 | 22 | AAAC84677 | Nucleotide sequenc |
| C 40 | 42.4 | 0.9 | 4466 | 21 | AAAI4663 | Nucleotide sequenc |
| C 41 | 42.4 | 0.9 | 4478 | 21 | AAAI4661 | Nucleotide sequenc |
| C 42 | 42.4 | 0.9 | 4547 | 21 | AAAI4664 | Nucleotide sequenc |
| C 43 | 42.4 | 0.9 | 4571 | 21 | AAAI4662 | Nucleotide sequenc |
| C 44 | 42.4 | 0.9 | 77536 | 21 | AAAI4651 | Nucleotide sequenc |
| C 45 | 42 | 0.8 | 203 | 19 | AAV30271 | Glutamine rich reg |

ALIGNMENTS

| | |
|--|----------------------|
| RESULT 1 | |
| AAV40485 | |
| ID AAV40485 standard; cDNA: 1302 BP. | |
| XX | |
| AC AAV40485; | |
| XX | |
| DT 09-NOV-1998 (first entry) | |
| XX | |
| DE Human secreted protein CO618_1 cDNA. | |
| XX | |
| KW Secreted protein: CO618_1; human; ds. | |
| XX | |
| OS Homo sapiens. | |
| XX | |
| FH Key | Location/Qualifiers |
| FT CDS | 123..1247 |
| FT | /*tag= a |
| XX | |
| PN W09831802-A1. | |
| XX | |
| PD 23-JUL-1998. | |
| XX | |
| PF 21-JAN-1998; 98WO-US01007. | |
| XX | |
| PR 20-JAN-1998; 98US-0010047. | |
| PR 21-JAN-1997; 97US-0072103. | |
| PR 16-JUN-1997; 97US-0877035. | |
| XX | |
| PA (GENY) GENETICS INST INC. | |
| XX | |
| PI Agostino MJ, Jacobs K, Lavallie ER, | Mccoy JM, Merberg D; |
| PI Racie LA, Spaulding V, Treacy M; | |
| XX | |
| DR WPI: 1998-414101/35. | |

SEE ID NO: 2
AC. NO. AAV40485
Database: N.-Geneseq-0601

DR P-PSDB; AAM29640.

XX New isolated polynucleotides and secreted proteins - obtained from
PT human foetal kidney, adult testes, placenta, adult brain and foetal
XX brain cDNA libraries

PS Claim 22; Page 71-72; 104pp; English.

XX This full-length cDNA clone, designated C0618.1 and deposited at
CC ATCC 98296, codes for a novel human secreted protein (see AAM29640).
CC It was isolated from a human adult brain cDNA library using
CC methods which are selective for cDNAs encoding or transmembrane proteins,
CC and was identified as encoding a secreted or transmembrane protein,
CC on the basis of computer analysis of the amino acid sequence of the
CC encoded protein. C0618.1 shows some similarity to human male bone
CC protein-encoding polynucleotides (see KIAA0237 protein. Novel secreted
CC identified in foetal kidney, adult testis, placenta, adult brain
CC and foetal brain cDNA libraries. These can be used in the
CC production of recombinant polypeptides in host (preferably
CC mammalian) cells. The polypeptides may have e.g. nutritional
CC activity, cytokine and cell proliferation/differentiation activity,
CC immunostimulant or immunosuppressive activity, haematopoiesis
CC regulating activity, tissue growth activity, activin/inhibin
CC activity, chemotactic or chemokinetic activity, haemostatic or
CC thrombolytic activity, receptor/ligand activity, antiinflammatory
CC activity, cadherin or tumour invasion suppressor activity, tumour
XX inhibition activity, or other activities.

Sequence 1302 BP; 404 A; 286 C; 322 G; 286 T; 4 other:

Query Match 17.0%; Score 844.8; DB 19; Length 1302;

Best Local Similarity 81.3%; Pred. No. 1.7e-231;

Matches 1058; Conservative 3; Mismatches 115; Indels 126; Gaps 2;

QY 3681 caggtatggatccacatagaggagcagatctgttccacaaatcccgagcagtgatg 3740
Db 1 caggtatggatccacatagaggagcagatctgttccacaaatcccgagcagtgatg 3740
QY 3741 taatgatgatctgcgggtttcaaggactagtgtcttcgtttccagagcacaact 3800
Db 61 taatgatgatctgcgggtttcaaggactagtgtcttcgtttccagagcacaact 3800
QY 3801 acatgtccctccacatcagagcggcggaggaagaacaggaatagtgctttacatca 3860
Db 121 acatgtccctccacatcagagcggcggaggaagaacaggaatagtgctttacatca 3860
QY 3861 aatgcaaaacagacagatggcgctgctgggggaagaactgtgacaaagaccagcatca 180
Db 181 aatgcaaaacagacagatggcgctgctgggggaagaactgtgacaaagaccagcatca 3920
QY 3921 gtggagacatgtgtcactgtgagaagaatgacggccggtccgactgtgagtgagcg 3980
Db 241 gtggagacatgtgtcactgtgagaagaatgacggccggtccgactgtgagtgagcg 3980
QY 3981 ccttggtaccagtgaggaagaagcggcgatctagcatgtgggccaatagtgagctatcg 4040
Db 301 ccttggtaccagtgaggaagaagcggcgatctagcatgtgggccaatagtgagctatcg 4040
QY 4041 ttgtgtctcagcgaagaatcgcaagtgtcctcactcagtcgcaaaacgaggaagagta 4100
Db 361 ttgtgtctcagcgaagaatcgcaagtgtcctcactcagtcgcaaaacgaggaagagta 4100
QY 4101 aaaaactcagcgaactttcaggaagaacagcagagaccgggttcacagtgagatgagga 4160
Db 421 aaaaactcagcgaactttcaggaagaacagcagagaccgggttcacagtgagatgagga 4160
QY 4161 actgtatgacccgacagccagcgagatcacagatgagcagatgtgacagctatagct 4220
Db 481 actgtatgacccgacagccagcgagatcacagatgagcagatgtgacagctatagct 4220
QY 4221 cggagaggaatctgtatcttccctgtgggtcgctgtgacctgtacagccagtgatgt 4280

Db 541 cagaagaaatctgtatcttccctgtgtctgtgtcgtgtgctctgtatgacagcttgatgt 600
QY 4281 tcttgatggtccttggtggtcctcagctagtgaggagcgcagacacctgtgactctcga 4340
Db 601 tcttgatggtccttggtggtcctcagctagtgaggagcgcagacacctgtgactctcga 660
QY 4341 tgggtgacatcagtgatggaatgtgataaagaaggaacagctgtgaggtgaaatcacc 4400
Db 661 tgggtgacatcagtgatggaatgtgataaagaaggaacagctgtgaggtgaaatcacc 720
QY 4401 gggggcggtcctgtgtgataaacaaggttcacaagacatgccaagcgtgataag 4460
Db 721 gggggcggtcctgtgtgataaacaaggttcacaagacatgccaagcgtgataag 4460
QY 4461 tgtatcgttagacaagagagctgcg----- 4486
Db 781 tgtatcattagatacggaggtctgcatgacaaaggagacagctgtgagtagaataca 840
QY 4487 ----- 4486
Db 841 tccgggcccgtggtctgtgtlaaacaaggttcacaagacatgccaagcgtgataag 900
QY 4487 -----atggccaaaagaacccaaggtgagcaaga 4517
Db 901 aagtgatctattagataaagcagaggtctgatatgccaagaagaagaagtgcaagaa 960
QY 4518 agacccttgagccctgtgtacacagcctctgtcctcgaagagagccccaagggagag 4577
Db 961 aagcccttgagccctgtgtacacagcctctgtcctcgaagagagccccaagggagag 4577
QY 4578 tgtacagatcatctgtcggggaagatgtgtgtatgatatgatacaaatcctttagggag 1020
Db 1021 tttacagatcatctgtcggggaagatgtgtgtatgatatgatacaaatcctttagggag 1080
QY 4638 tggcccaatactctttagatgaactggaactgaactcaacaatgtgattggtgtcaaac 4657
Db 1081 tggcccaatactctttagatgaactggaactgaactcaacaatgtgattggtgtcaaac 4657
QY 4698 tcttcacctctctcctccttagatgataccaacctgtgacacctgtgacaagaagctccc 4757
Db 1141 tttccacctctctcctccttagatgataccaacctgtgacacctgtgacaagaagctccc 1200
QY 4758 aatgtctcttggaagaagtcttaccagaccttctactctgttcatagcaact---ataa 4814
Db 1201 aatcatctcttggaagaagtcttaccagaccttctactctgttcatagcaact---ataa 1260
QY 4815 actgtgtcacaacacagcagatgatacaaaaacagagaaga 4856
Db 1261 atgtgtgtcacaacacagcagatgatacaaaaacagagaaga 1302

RESULT 2

ID AAC60011 standard; cDNA; 2276 BP.

AC AAC60011;

DT 26-JAN-2001 (first entry)

DE Human secreted protein gene 46 SPQ ID NO:56.

KW Human; secreted protein; diagnosis; cyrostatic; immunosuppressive;
KW neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiac; gene therapy;
KW neurological disease; cardiovascular disorder; wound healing;
KW chromosome 6; ss.
XX
OS Homo sapiens.
XX
PN WO200058356-A1.

| | | |
|----|------|--|
| PD | XX | 05-OCT-2000. |
| XP | PF | 22-MAR-2000; 2000WO-US07535. |
| XX | XX | 26-MAR-1999; 99US-0126511. |
| PR | XX | 17-DEC-1999; 99US-0172413. |
| XX | PA | (HUMA-) HUMAN GENOME SCI INC. |
| XX | PI | Rosen CA, Ruben SM, Komatsoulis G; |
| DR | XX | WPI: 2000-594639/56. |
| XX | DR | P-PSDB: AAB34818. |
| PT | CC | Fifty nucleic acid molecules encoding human secreted proteins, useful in the prevention, treatment and diagnosis of cancer, immune disorders, cardiovascular disorders and neurological diseases - |
| PS | CC | Claim 1; Page 372-373; 425pp; English. |
| XX | CC | The polynucleotide sequences given in AAC595966 to AAC60015 encode the human secreted proteins given in AAB34773 to AAB34822. AAB34823 to AAB34852 represent human secreted polypeptide sequences and proteins homologous to them, which are given in the exemplification of the present invention. Human secreted proteins have activities based on the tissues and cells the genes are expressed in. Examples of activities include: cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor; vulnerable; anticoagulant; antibacterial; antifungal; antiparasitic; and cardiac. The polynucleotides and polypeptides are useful for preventing, treating or ameliorating a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. The polynucleotides are useful for chromosome identification. They are also useful as probes for diagnosing a disorder related to the female reproductive system, particularly breast and/or ovary cancer. They are also useful in the gene therapy of breast and ovarian cancer. Nucleic acids, protein, antibodies, agonists and antagonists from the present invention are useful in the diagnosis, treatment and prevention of cancer, immune disorders, cardiovascular disorders, wound healing, neurological diseases and infectious disease. AAC59957 to AAC59965 and AAB34772 represents sequence used in the exemplification of the present invention. |
| XX | CC | Sequence 2276 BP; 699 A; 451 C; 436 G; 689 T; 1 other: |
| SQ | XX | |
| | | Query Match 7.68; Score 377; DB 21; Length 2276; |
| | | Best Local Similarity 71.9%; Pred. No. 2.6e-97; |
| | | Matches 507; Conservative 0; Mismatches 195; Indels 3; Gaps 1; |
| OY | 4100 | aaaagaactgcggacgtctcagaagaagccaggaccggcgctcagtcagtgaatgag 4159 |
| Dd | 96 | aaaaagtttaaaagaattcaatccacgaagaagacaagaacaagcaatgcgtgaatatgaga 155 |
| OY | 4160 | aactgcgatgcccgccagccagccggagataccaagatacagatgacagcatagc 4219 |
| Dd | 156 | aa--gattgagtagaacgacgcgagccgagagctactatgtagtgagatcaacaacgttaacgc 212 |
| OY | 4220 | tccgaagaagaatctgattctcccttgaggctccgcctgcgctctgacagccagttcagtat 4279 |
| Dd | 213 | tctagaggaacaatttaattctcttgagtgctgcagtcggagcgtgacagttcaattcagttgat 272 |
| OY | 4280 | tctcgtgatggccctgggcccctgcataagtcagtcaggacccagacccttgctctctcgca 4339 |
| Dd | 273 | tcttcgtatgtagttatgggacccagccagcgtctgtgctgcgcaccaaacctctgcccccttgca 332 |
| OY | 4340 | atgggtacatctcaggtctgagatgtagtataaaaaggagacagcttgaggtagaataatc 4399 |
| YY | 333 | atgggtatataacaatagatgtagtggagcaaaaaggcccaattggaagtgtgaatcattc 392 |
| OY | 4400 | cgggcgcgcgacctgtgtgtataaaaccaggttccaagacatctgcagcacccgtatgtcaag 4459 |

| | | | |
|----------|---|--|------|
| Db | 393 | agagcagcgaagccctcacacaagaagccggtgtcccaatctaacctgtctccatagtcaaa | 452 |
| Qy | 4460 | gtgtactcgttaagacaacggaagtcgtcgaagcacaagaacaaagttgcgagaag | 4519 |
| Db | 453 | gtatattctttggaaaatggggcctgtatagccaagaagaacagaattgcacgaata | 512 |
| Qy | 4520 | acccttgagagcccccgtaccagacgtctgtctccttcgagagagccccaaggaggtg | 4579 |
| Db | 513 | acccttgatcccttgttaccagacgtctgttctgttttttgtaagttccacagtgtaagt | 572 |
| Qy | 4580 | ttacagatcatgtctcgaggagatattgctgtatgtatgtatccaatccctttatggagt | 4639 |
| Db | 573 | cttcagtgatgttgcctcgggagactatggcagaatgtgcccacaattgttaatgggtg | 632 |
| Qy | 4640 | gcccaactactcttagatgaactgtgaactacacacatgtgtatgtatgtttcaatc | 4699 |
| Db | 633 | gtcagatcatctgttggaagaactggaactggaactgtgcagcatgtatcgtatgtataaat | 692 |
| Qy | 4700 | ttccctctctctccctcagtagatccacaactcctcggaactctgcagaagaagcttccaa | 4759 |
| Db | 693 | ttccacacgctctactatgtgtatcccaactactcctccacaccggcggttccag | 752 |
| Qy | 4760 | tcgctctcggaaagtctacaggacgtctctactcctgttatag | 4804 |
| Db | 753 | tcactctcgaaagttcaactggcctccctgtattcgatcatag | 797 |
| RESULT 3 | | | |
| ID | AAV90372/c | | |
| XX | AAV90372 standard; cDNA; 355 BP. | | |
| AC | AAV90372; | | |
| XX | | | |
| DT | 15-FEB-1999 (first entry) | | |
| XX | | | |
| DE | EST clone DM118. | | |
| XX | | | |
| KM | Human; secreted protein; expressed sequence tag; EST; haematopoiesis; | | |
| KM | tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic; | | |
| KM | receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour; | | |
| KM | gene therapy; ss. | | |
| XX | | | |
| OS | Homo sapiens. | | |
| XX | | | |
| PN | WO9845436-A2. | | |
| XX | | | |
| PD | 15-OCT-1998. | | |
| XX | | | |
| PF | 10-APR-1998; 98WO-US06955. | | |
| XX | | | |
| PR | 10-APR-1997; 97US-0838821. | | |
| XX | | | |
| PA | (GENM) GENETICS INST INC. | | |
| PI | Agostino MJ, Jacobs K, Lavallie EF, McCoy JM, Merberg D; | | |
| PI | Racie LA, Spaulding V, Treacy M; | | |
| XX | | | |
| DR | WPI; 1999-070077/06. | | |
| XX | | | |
| PT | New polynucleotides encoding human secreted proteins - derived from | | |
| PT | e.g. human blood, kidney, foetal lung, placenta, testes, brain, | | |
| PT | ovary, pituitary, retina and colon cDNA libraries. | | |
| XX | | | |
| PS | Claim 1; Page 519-520; 618bp; English. | | |
| CC | The present sequence represents a human expressed sequence tag (EST). | | |
| CC | The polynucleotide, which is a secreted EST, and the encoded protein | | |
| CC | are predicted to have useful biological activities which would make | | |
| CC | them suitable for treating, preventing or ameliorating medical | | |
| CC | conditions in humans and animals, although no supporting data is | | |
| CC | given. Suggested activities include nutritional activity, immune | | |
| CC | stimulating or suppressing activity, haematopoiesis regulating | | |

CC activity, tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
 CC activity, receptor/ligand activity, anti-inflammatory activity,
 CC adherin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The polynucleotide may also be useful for gene therapy.
 XX
 XX Sequence 355 BP; 93 A; 82 C; 83 G; 97 T; 0 other;

Query Match 4.2%; Score 207.6; DB 20; Length 355;
 Best Local Similarity 90.2%; Pred. No. 2.9e-49;
 Matches 222; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3844 cagtgcttaccatccaaatgcaaaacagacagatggcggtgagggaacttgac 3903
 Db 247 CAGTGCTTTACATCCAAATGCAAGCAGACAAATGGCGCATATCAGGAAGAACATGAC 188
 QY 3904 caaagaccagcagcagtcagtgagacatgtctcactggaagaatgacgagcagc 3963
 Db 187 AAAAAGCAGCCAGCATCATGTGGAGACATGTGCTCACTGGAGAAATGATGCGACCACTC 128
 QY 3964 cgacactgcagtgaggcgccctgggtacacagtgagcaagaagcgagatcagcaltgggc 4023
 Db 127 TGACATCGCAGTGGCGCAGCTTGGCCACCAAGTGGCGCTTACGCTTGGTGC 68
 QY 4024 caaatgtagctatgttggtctctcaggaagaatgcagtgccctcactcaactcagcca 4083
 Db 67 CAAATGCTAGCTATGCTGTGCTGTCACGAAAGTCGAGTGTCTCAGCTCAACGCG 8
 QY 4084 aaccga 4089
 Db 7 ACTCGA 2

RESULT 4
 ID AAV89630
 XX AAV89630 standard; cDNA; 355 BP.
 AC AAV89630;

DT 15-FEB-1999 (first entry)
 XX
 DE EST clone CO618.
 XX

KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
 KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
 KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
 KW gene therapy; ss.

XX Homo sapiens.
 OS

XX WO9845436-A2.
 PN

XX 15-OCT-1998.
 PD

XX 10-APR-1998; 98WO-US06955.
 PF

XX 10-APR-1997; 97US-0838821.
 PR

XX (GEMV) GENETICS INST INC.
 PA

XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 PI Racie LA, Spaulding V, Treacy M;
 DR WPI; 1999-070077/06.
 XX

XX New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries.
 XX

PS Claim 1; Page 279; 618pp; English.
 XX

CC The present sequence represents a human expressed sequence tag (EST).

CC The polynucleotide, which is a secreted EST, and the encoded protein
 CC are predicted to have useful biological activities which would make
 CC them suitable for treating, preventing or ameliorating medical
 CC conditions in humans and animals, although no supporting medical
 CC given. Suggested activities include nutritional activity, immune
 CC stimulating or suppressing activity, haematopoiesis regulating
 CC activity, tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
 CC activity, receptor/ligand activity, anti-inflammatory activity,
 CC cadherin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The polynucleotide may also be useful for gene therapy.
 XX
 XX Sequence 355 BP; 97 A; 83 C; 82 G; 93 T; 0 other;

Query Match 4.2%; Score 207.6; DB 20; Length 355;
 Best Local Similarity 90.2%; Pred. No. 2.9e-49;
 Matches 222; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 3844 cagtgcttaccatccaaatgcaaaacagacagatggcggtcggggaacttgac 3903
 Db 109 cagtgcttaccatccaaatgcaaaacagacagatggcggtcggggaacttgac 168
 QY 3904 caaagaccagcagcagtcagtgagacatgtctcactggaagaatgacgagcagc 3963
 Db 169 aaaaagaccagcagcagtcagtgagacatgtctcactggaagaatgacgagcagc 228
 QY 3964 cgacactgcagtgaggcgccctgggtacacagtgagcaagaagcgagatcagcaltgggc 4023
 Db 229 tgacactgcagtgaggcgccctgggtacacagtgagcaagaagcgagatcagcaltgggc 288
 QY 4024 caaatgtagctatgttggtctctcaggaagaatgcagtgccctcactcaactcagcca 4083
 Db 289 caaatgtagctatgttggtctctcaggaagaatgcagtgccctcactcaactcagcca 348
 QY 4084 aaccga 4089
 Db 349 actcga 354

RESULT 5
 ID AAC74934
 XX AAC74934 standard; cDNA; 378 BP.
 AC AAC74934;

DT 08-FEB-2001 (first entry)
 XX

XX Human ORFX ORF489 polynucleotide sequence SEQ ID NO:977.
 DE

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnerability; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCIP; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.

XX Homo sapiens.
 OS

XX WO200058473-A2.
 PN

XX 05-OCT-2000.
 PD

XX 31-MAR-2000; 2000WO-US08621.
 PF

[illegible]

CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 1.9%; Score 93.4; DB 22; Length 936;
Best Local Similarity 4.2%; Pred. No. 2.8e-16;
Matches 34; Conservative 455; Mismatches 336; Indels 0; Gaps 0;

Qy 2099 ttaaggtctgtagaggaagaatgcacgtcacttctgcatattaccaca 215
::: : : : : : : : : : : : : : : : : :
Db 42 ww 101

Qy 2159 gtaaaaaaggaagttagctgatacgtlagacatcttagaccagtgatgaacttg 221
::: :: : : : : : : : : : :
Db 102 www 161

QY 2219 gaatggaatcggagagctatttgcgaaggagccacatttgagaagtltacaacatatcta 227
:: :: : :::: : : : : : : : : : : : :
Db 162 www 221

OY 2279 gaatccaaacctgaaoccaagttgtgacttggtttcgaagcgaattgagatatccct 233
::
Db 222 www 281

QY 2359 agaatacctgtagcagcatgcacactgtgaatccagtcttcagtclcatlttgaaclcaa 2399
 : : : : : : : : : : : : : : : : : :
Db 282 wwwwwwwwwCwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 341

QY 239 aaatgacgcgtccttctatataccgttaacctcaacccatgagtcctgcatgctgaagat 245E
 Db 342 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgctlaawwww 401

```

0Y 2459 gtcccgagttctatctgacagcttccaataaactatggttgacaagttgycac 2518
      :  :::::  :  :: :::::  ::  ::  :  ::  :
Db 402 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 461

```

| | | | | | | |
|----|------|-------------------|------------------|------------------|--------------|------|
| QY | 2519 | cgattgacggttaacaa | ttttgagagcaagaag | atctccctccaaggaa | gatggagagcca | 2578 |
| | | : : : : : | : : : : | : : : : | : : : : | |
| Db | 462 | www | www | www | www | 521 |

```
QY 2579 aggaatcctatgttaagaattactccttcagatagaatgataaaaataagaaga 26388
      : :: :: :: :: :: : : :: : : :: :: :: : : :
Db 522 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 581
```

```

QY 2639 acataaacagtcagaagaaacttggaaacccaatgtaaccgaacttcatttctcct 26388
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 582 wwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 641

```

[illegible]

Dy 2159 gttcgaagaagagcgcaattcttagsgagatlttaattgaaacacgttgg 2810
:: : : : : : : : : : : : : : : :
Db 702 ww 761

qy 2819 ccagatgatgagccgcacgtgtataaacgtcagaccocatgtcttcctcatggccactc 2878
:: :: :: : :: :: ||| | ||| | ||| |
Db 762 wwwwwwwwwwwwwwwwwwwwwwwwggccccccccccccccccccccccccccc 821

```

QY 28/9 cctgcgccttccccatactgccc 2903
    ||| ||| ||| | |||
Db 822 cccccccccccccccccccc 846

```

RESULT 8
AAF58257

RESULT 8
AAF58257

| | | |
|---|--|------------------------|
| ID | AAE58257 | standard; DNA; 936 BP. |
| XX | | |
| AC | AAE58257; | |
| XX | | |
| DT | 24-APR-2001 | (first entry) |
| XX | | |
| DE | Oligonucleotide D1954. | |
| XX | | |
| KM | Electron-transfer group; ETM; mismatch; genotyping; | |
| KM | gene expression; ss. | |
| OS | Synthetic. | |
| XX | | |
| PN | WO200107665-A2. | |
| XX | | |
| PD | 01-FEB-2001. | |
| XX | | |
| PF | 26-JUL-2000; 2000MO-US20476. | |
| XX | | |
| PR | 26-JUL-1999; 99US-0145695. | |
| PR | 17-MAR-2000; 2000US-0190259. | |
| XX | | |
| PA | (CLIN-) CLINICAL MICRO SENSORS INC. | |
| XX | | |
| PI | Unex RM; | |
| XX | | |
| DR | WPI: 2001-159728/16. | |
| XX | | |
| PT | Nucleic acids containing electron-transfer group, useful as labels in | |
| PT | hybridization assays, e.g. for genotyping, allowing repeat analyses on | |
| XX | a single surface | |
| PS | | |
| XX | Example 6; Page 127; 159pp; English. | |
| CC | The present invention relates to a composition comprising two nucleic | |
| CC | acids each containing an electron-transfer group (ETM) having | |
| CC | different redox potentials. The invention is used for electronic | |
| CC | detection of nucleic acids, especially of substitutions (mismatches) | |
| CC | and single-nucleotide polymorphisms, e.g. for genotyping, | |
| CC | monitoring gene expression. | |
| XX | | |
| SO | Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other; | |
| <hr/> | | |
| Query Match | 1.9%; Score 93.4; DB 22; Length 936; | |
| Best Local Similarity | 4.2%; Pred. No. 2.8e-16; | |
| Matches 34; Conservative 435; Mismatches 330; Indels 0; Gaps 0; | | |
| QY 2059 | ttaaaggttgtaggagaagaatgaatcagtcagctgtgcatttaccaca | 2158 |
| Dd 42 | www..... | 101 |
| QY 2159 | gtaaaaaaggaaagttagctgatctgtaggaaacttagaccagtgatgaagctctg | 2218 |
| Dd 102 | www..... | 161 |
| QY 2219 | gaatggaatggaggacttatcgcaaggagccacctttggaggaaatttaaacatttcta | 2278 |
| Dd 162 | www..... | 221 |
| QY 2279 | gaatccaacctgaaccaagtttagcttgttttcaaggccaattggagataatcct | 2338 |
| Dd 222 | www..... | 281 |
| QY 2339 | agaatactgtatagcagcatgacaaactggaatccagttctagctcaattgaatccta | 2398 |
| Dd 282 | www..... | 341 |
| QY 2399 | aaaatggacgcgtccctctatatatccgttacccaccatgagtcctgcatgctggagat | 2458 |
| Dd 342 | www..... | 401 |
| QY 2459 | gtcccgcagttcttactctgcacagcttccaataaactatggttgacaaggttgatcac | 2518 |

| | | | |
|----|------|--|------|
| Db | 402 | | 461 |
| Qy | 2519 | cagttgatagttacaaatttggagcaaaagatctctccagggaaga tggagagcca | 2578 |
| Db | 462 | | 521 |
| Qy | 2579 | aggaaatcctatgttaagaatttactctccagatagaagtgataaaaa laagagaga | 2638 |
| Db | 522 | | 581 |
| Qy | 2639 | acaaaacagtcaagaaaacttggaaacccaatggaaccagacttacttctct | 2698 |
| Db | 582 | | 641 |
| Qy | 2699 | gtccaccggaagaatctcgtgaacgaatgctggaattacctttggatacagctaga | 2758 |
| Db | 642 | | 701 |
| Qy | 2759 | gtctgagaagaagaagagcaatcttagagagatttaattgaatggaacagcttg | 2818 |
| Db | 702 | | 761 |
| Qy | 2819 | ctgatatgtagccgcaactggtataagcttcagaaccatgattctcctaatgacatc | 2878 |
| Db | 762 | | 821 |
| Qy | 2879 | cctgcgccttcccacatctgccc 2903 | |
| Db | 822 | cccccccccccccccccccccccc 846 | |

| | |
|----------|--|
| RESULT | 9 |
| AAFS8259 | |
| ID | AAFS8259 standard; DNA; 936 BP. |
| XX | |
| AC | AAFS8259; |
| XX | |
| DT | 24-APR-2001 (first entry) |
| XX | |
| DE | Oligonucleotide D2004. |
| XX | |
| KW | Electron-transfer group; ETM; mismatch; genotyping; |
| KW | gene expression; ss. |
| XX | |
| OS | Synthetic. |
| XX | |
| PN | WO200107665-A2. |
| XX | |
| PD | 01-FEB-2001. |
| XX | |
| PF | 26-JUL-2000; 2000WO-US20476. |
| XX | |
| PR | 26-JUL-1999; 99US-0145695. |
| PR | 17-MAR-2000; 2000US-0190259. |
| XX | |
| PA | (CLIN-) CLINICAL MICRO SENSORS INC. |
| XX | |
| UM | Umek RM; |
| PI | |
| DR | WPI; 2001-159728/16. |
| XX | |
| XX | |
| PT | Nucleic acids containing electron-transfer group, useful as labels in |
| PT | hybridization assays; e.g. for genotyping, allowing repeat analyses on |
| XX | a single surface - |
| XX | |
| PS | |
| XX | Example 6; Page 128; 1599p; English. |
| XX | |
| CC | The present invention relates to a composition comprising two nucleic |
| CC | acids each containing an electron-transfer group (ETM) having |
| CC | different redox potentials. The invention is used for electronic |
| CC | detection of nucleic acids, especially of substitutions (mismatches) |
| CC | and single-nucleotide polymorphisms, e.g. for genotyping, |

CC monitoring gene expression.

SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

| Query Match | 1.98; Score 93.4; DB 22; Length 936; |
|-------------|--------------------------------------|
|-------------|--------------------------------------|

Best Local Similarity 4.28; Pred. No. 2.8e-16;
Matches 34; Conservative 435; Mismatches 336; Indels 0; Gaps 0;

| | | | |
|----|------|---|------|
| QY | 2099 | ttaaagtctgtagaagaagatgatgatacggcttcgttgcatattatacca | 2158 |
| Db | 42 | | 101 |
| QY | 2159 | gtaaaaaagaagtctatgctga tactgtaggacatcttagaccagtgatgaactctg | 2218 |
| Db | 102 | | 161 |
| QY | 2219 | gaatggaatvggagcgcatctgcagaagaccacattgaggaagttacaacattacta | 2278 |
| Db | 162 | | 221 |
| QY | 2279 | gaatccaacctgaaccacagcttgagctgtgttttaagcgcaactggagatctct | 2338 |
| Db | 222 | | 281 |
| QY | 2339 | agaatacctgatagcacgcacatgcacaactggaatccagttctagctattgaaatcoa | 2398 |
| Db | 282 | | 341 |
| QY | 2399 | aaatggaacgctctctatatactcgttactccaatgagtcctcgcatctgaggat | 2458 |
| Db | 342 | | 401 |
| QY | 2459 | gtcccgagttctatctatcgtgacagcttcaataaactatggttgacaagttgttca | 2518 |
| Db | 402 | | 461 |
| QY | 2519 | caattgatagtacaattctggagacaagatctcctccaaggaaatvggagcca | 2578 |
| Db | 462 | | 521 |
| QY | 2579 | aggatctctatgttaagatttactctctccaagatagaagtgataaaaaataagaaga | 2638 |
| Db | 522 | | 581 |
| QY | 2639 | acaaaacagtcacaagaacatttggaaccacaatggaacagacttcatlactct | 2698 |
| Db | 582 | | 641 |
| QY | 2699 | gtccacccgaagaagaatcogtgaagaatgctggaataacccttgatcaactaga | 2758 |
| Db | 642 | | 701 |
| QY | 2759 | gttcgagaagaagagcgaaatcttagagagatttaattgaattgaaacagcttg | 2818 |
| Db | 702 | | 761 |
| QY | 2819 | ctagatgatgacgcgacctggtataaagctgcagaccacatgatgtctccatctgacac | 2878 |
| Db | 762 | | 821 |
| QY | 2879 | ccctgcacctcccatatctgcacc | 2903 |
| Db | 822 | | 846 |

RESULT 10

ID AAF58262 standard; DNA; 936 BP.

AC AAF58262;

DT 24-APR-2001 (first entry)

XX
DE Oligonucleotide D2007

KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.

OS Synthetic.

PN W0200107665-A2.

PD 01-FEB-2001.

PF 26-JUL-2000;

PR 26-JUL-1999

XX

XX

XX

XX

PT hybridization

XX

XX

CC acids each c

CC detection of

CC monitoring &

Sequence 936

1

Best Local Simi

○
○
○
○
—
—

25

0
2
7
0

| Query Match | Similarity | 1.9% | Score 93.4 | DB 22 | Length 936 |
|-------------|------------|--|------------------|------------|------------|
| Best Local | Similarity | 4.2% | Pred No. 2.8e-16 | | |
| Matches | 34 | Conservative | 435 | Mismatches | 336 |
| | | | | Indels | 0 |
| | | | | Gaps | 0 |
| 0Y | 2099 | ttaagagttgttgaggagaagatgagctgaatcaagctgcagcttgttcattattaccocaa | 2158 | | |
| | 42 | ##### | 101 | | |
| 0Y | 2159 | gtataaaagggaagtttcgtctgatactgtaggacactttagaccagtgatgaagctctg | 2218 | | |
| | 102 | ##### | 161 | | |
| 0Y | 2219 | gaatgaatggagagctatcgaaaggacacatttggaggagtttaacacatttcta | 2278 | | |
| | 162 | ##### | 221 | | |
| 0Y | 2279 | gaatccaaacctgaaccacagttgagcttgttccaaggccaattggagatctcct | 2338 | | |
| | 222 | ##### | 281 | | |
| 0Y | 2339 | agaatacctgatagcagcaatgacaaactgtaatcagttctbaggctatttgaatctcaa | 2398 | | |
| | 282 | ##### | 341 | | |
| 0Y | 2399 | aaatggagccgtccctctatatctcgtttaccaccaaccatgagtccctcgatctgaggat | 2458 | | |
| | 342 | ##### | 401 | | |
| 0Y | 2459 | gtcccgagttcttatctcggacaagcttcaataaaacatggtttgacaaggttgtaac | 2518 | | |
| | 402 | ##### | 461 | | |
| 0Y | 2519 | caatgatagttacacatttggagcaaaaggatcccttcacaggagaatcgaggagcca | 2578 | | |

```
Db 462 www..... 521
Oy 2579 aggaatccttaagtaagattccttcctccaagatagaagtataaaataagagaaga 2638
Db 522 www..... 581
Oy 2639 acaaaaacagtaagaacttctggaacccaatggaacagacttcatattctcct 2698
Db 582 www..... 641
Oy 2699 gtccacccgaagaaatccgtgaacgaatctggaatcttgccttggatcaagctaga 2758
Db 642 www..... 701
Oy 2759 gtccgagaagaagagcgaattcttaagagagatttaattgaattggaacagcttg 2818
Db 702 www..... 761
Oy 2819 ctatagtatgagcgcactggtataagctgcagaccatgatgtctcctcattgcaactc 2878
Db 762 www..... 821
Oy 2879 cctgcacctcccatatctgccc 2903
Db 822 cccccccccccccccccccc 846

RESULT 11
AAF58255
ID AAF58255 standard; DNA; 938 BP.
AC AAF58255;
XX 24-APR-2001 (first entry)
XX Oligonucleotide D1876.
DE Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX Synthetic.
OS MO200107665-A2.
PN 01-FEB-2001.
PD 26-JUL-2000; 2000WO-US20476.
PE 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX (CLIN-) CLINICAL MICRO SENSORS INC.
PA Umek RM;
PI WPI: 2001-159728/16.
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX Example 6; Page 127; 159pp; English.
XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
XX Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;
SO
```

```
Query Match 1.9%; Score 93.4; DB 22; Length 938;
Best Local Similarity 4.2%; Pred. No. 2,8e-16;
Matches 34; Conservative 435; Mismatches 336; Indels 0; Gaps 0;

Oy 2099 tlaaagttgtaggaaagabtaagtaacagtcagtcgacttgttgatattaccaaa 2158
Db 42 www..... 101
Oy 2159 gtaaaaaaaggaagttagctgatactgtagacacacttaagacaagtgatgaagcttg 2218
Db 102 www..... 161
Oy 2219 gaatggaatggaggactatgcaaggagccacattgaggaaagttaacaattattcta 2278
Db 162 www..... 221
Oy 2279 gaatccaaacctgaaacaaagtgtgagctgtgttccaaggccaattggaatattcct 2338
Db 222 www..... 281
Oy 2339 agaatcaacctgatagcagcatgcacaaactggaatccagttctgacttaattgaatccaa 2398
Db 282 www..... 341
Oy 2399 aaaaatgaccgtcctctctatccgttaacctcaaccatgagtgactgcatgtagggat 2458
Db 342 www..... 401
Oy 2459 gtccgcagttcttattctgacagcttccaataaactatggttgaacaggtgtgtcac 2518
Db 402 www..... 461
Oy 2519 cagtgatagttacaattttgtagagcaaggaatcctcccgaggagaagtggagagca 2578
Db 462 www..... 521
Oy 2579 aggaatccttaagtaagattccttcctccagatagaagtataaaataagagaaga 2638
Db 522 www..... 581
Oy 2639 acaaaaacagtaagaanaaacttggaaacccaatggaacagacttcatattctcct 2698
Db 582 www..... 641
Oy 2699 gtccacccgaagaaatccgtgaacgaatgctggaatccttggatcaagctaga 2758
Db 642 www..... 701
Oy 2759 gtccgagaagaagagcgaattcttaagagagatttaattgaattggaacagcttg 2818
Db 702 www..... 761
Oy 2819 ctatagtatgagcgcactggtataagctgcagaccatgatgtctcctcattgcaactc 2878
Db 762 www..... 821
Oy 2879 cctgcacctcccatatctgccc 2903
Db 822 cccccccccccccccccccc 846

RESULT 12
AAF58252/c
ID AAF58252 standard; DNA; 936 BP.
AC AAF58252;
XX 24-APR-2001 (first entry)
XX Oligonucleotide D1835.
DE Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
```

XX Synthetic.
OS
XX
PN WO200107665-A2.
XX
XX 01-FEB-2001.
XX
XX 26-JUL-2000; 2000WO-US20476.
XX
XX 26-JUL-1999; 98US-0145695.
XX 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX Umek RM;
XX
XX
XX WPI; 2001-159728/16.
XX
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
XX
XX Example 6; Page 127; 159pp; English.
XX
XX
XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
XX
XX Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

| | | | |
|--------|---------------------------------|--|------|
| OY | 2570 | gggaagcCcaagaatctctaTgttaagattacttcccttcagataagaagtgtataaaat | 2625 |
| Db | 252 | MM | 193 |
| OY | 2630 | aagagaagaacaaaacagTcaaagaaccttgaaaccacaatgaaacagacttcat | 2689 |
| Db | 192 | MM | 133 |
| OY | 2690 | tattctctgtccacCGaagaatccgt'gaacgaatgcTgaaattaaccttggat | 2749 |
| Db | 132 | MM | 73 |
| OY | 2750 | caagctagaagtlcgagaagaagagcgaaTtcttagagagatttaattgaattggaa | 2809 |
| Db | 72 | MM | 13 |
| OY | 2810 | acagacttt 2817 | |
| Db | 12 | MMMMMMMMMM 5 | |
| RESULT | 13 | | |
| ID | AAF58254/C | | |
| xx | AAF58254 standard; DNA; 936 BP. | | |
| AC | AAF58254; | | |
| XX | | | |
| DT | 24-APR-2001 (first entry) | | |
| XX | | | |
| DE | Oligonucleotide D1875. | | |
| XX | | | |

| | Query Match | 1.98, Score 92.8, DB 22, Length 936, Best Local Similarity 0.9%; Pred. No. 4, 2e-16; Matches 7; Conservative 462; Mismatches 319; Indels 0; Gaps 0; |
|----|---|---|
| OY | gctcgatatttataaagaagcgtttaagaagatgagagtgtagtaccctcgagaccagagaca | 2089 |
| DB | GGGGG | 733 |
| OY | atgctggagcttaaaagttgttgaaggaaagatagctgaatcagtcgacattgttcatt | 2149 |
| DB | 732 | 673 |
| OY | attcccaagcaaaaaaaggaagtttagctgtatctgtagacatcttagaccagtgat | 2209 |
| DB | 672 | 613 |
| OY | gaagcttggaatggaatggagagctattgcaagagcgacattggaagattacaac | 2269 |
| DB | 612 | 553 |
| OY | ataatcttagaatccaaacctgaaaccaagttgagcttgttcttcaaggccaattga | 2329 |
| DB | 552 | 493 |
| OY | gatatctctagaatacctgatagacgacatgacaaactggaatccagttctagctcatt | 2389 |
| DB | 492 | 433 |
| OY | gaatctcaaaaaaggaacgctcctctatctacgttacctaccacatgagtcctgcatg | 2449 |
| DB | 432 | 373 |
| OY | ctgaggatgtcccgacgtttcttactggaacgcttcaataaactatggttacaag | 2509 |
| DB | 372 | 313 |
| OY | gttggtaccacgattgataagttacaatttggagcaaaagatcctccctccaggaagat | 2569 |
| DB | 312 | 253 |

XX gene expression; ss.
 OS Synthetic.
 PN WO200107665-A2.
 XX
 PD 01-FEB-2001.
 XX
 PF 26-JUL-2000; 2000WO-US20476.
 XX
 PR 26-JUL-1999; 99US-0145695.
 PR 17-MAR-2000; 2000US-0190259.
 XX
 PA (CLIN-) CLINICAL MICRO SENSORS INC.
 XX
 PI Umek RW;
 XX
 DR WPI; 2001-159728/16.
 XX
 PT Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface
 XX
 PS Example 6; Page 127; 159pp; English.
 XX
 CC The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ERM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.
 XX
 SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

```
11 .....
Db 792 GGGGCG..... 733
Qy 2090 atgcggccttaaggttgtagaagaagatgactgaatcagtgacttctgcatc 2149
Db 732 ..... 673
Qy 2150 attaccaagtaaaaaagaagtttagctgatactgtgagacatttagaccagtgat 2209
Db 672 ..... 613
Qy 2210 gaagcttggaatgggaatggagagctatgcagaagaccacatttgagaagtttaac 2269
Db 612 ..... 553
Qy 2270 attattcagaatccaacctgaaccacaagttgagctgtgtttccaaggccaattgga 2329
Db 552 ..... 493
Qy 2330 gatattcctagatacctgatagcacgcatgcacaactggaatccagttctcatctt 2389
Db 492 ..... 433
Qy 2390 gaatctcaaaaaatggaccgtctctctatatacgcgttaacctcaaccatgagctcgcgatg 2449
Db 432 ..... 373
Qy 2450 ctgagagatgcccgagttcttactcgcagcagcttcaataaactatggttgacaag 2509
Db 372 ..... 313
Qy 2510 gtgtgtccagttgatagttacaatttggagacaagatccctccagggaagat 2569
Db 312 ..... 253
Qy 2570 gggagggccaagatcctatgttaagattactctctccagatagaagtgaataaat 2629
Db 252 ..... 193
Qy 2630 aagagaagaacaaaaacagtcagaagaacttggaaacccaatggaacagcagacttcat 2689
Db 192 ..... 133
Qy 2690 tattctctgtccaccggaagaatccgtgaacgaatgctggaatattacccttggat 2749
Db 132 ..... 73
Qy 2750 caagctagatcgaagaagaagcgaattctcttagagagatttaattgaattgaa 2809
Db 72 ..... 13
Qy 2810 acaagctt 2817
Db 12 ..... 5
Db 12 ..... 5

RESULT: 14
ID AAF58257/c
ID AAF58257 standard; DNA; 936 BP.
AC AAF58257;
XX
XX 24-APR-2001 (first entry)
XX
XX Oligonucleotide D1954.
XX
XX Electron-transfer group; ETM; mismatch; genotyping;
XX gene expression; ss.
XX
XX Synthetic.
XX
XX MO200107665-A2.
XX
XX
```

```
PD 01-FEB-2001.
XX
XX 26-JUL-2000; 2000MO-US20476.
XX
XX 26-JUL-1999; 99US-0145695.
XX
XX 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX
XX umek RM;
XX
XX WPI; 2001-159728/16.
XX
XX Nucleic acids containing electron-transfer group (ETM) having
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface
XX
XX Example 6; Page 127; 159pp; English.
XX
XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
XX
XX Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other:
XX
XX
XX Query Match 1.9%; Score 92.8; DB 22; Length 936;
XX Best Local Similarity 0.9%; Pred. NO. 4,2e-16;
XX Matches 7; Conservative 462; Mismatches 319; Indels 0; Gaps 0;
XX
XX
XX Qy 2030 gtcgatttattaaataagcgtttaaagatggagtgctacctgagactcaggaga 2089
XX 11 ..... 733
XX Db 792 GGGGCG..... 733
XX
XX Qy 2090 atgcggccttaaggttgtagaagaagatgactgaatcagtgacttctgcatc 2149
XX .. ..... 673
XX Db 732 ..... 673
XX
XX Qy 2150 attaccaagtaaaaaagaagtttagctgatactgtgagacatttagaccagtgat 2209
XX ..... 613
XX Db 672 ..... 613
XX
XX Qy 2210 gaagcttggaatgggaatggagagctatgcagaagaccacatttgagaagtttaac 2269
XX ..... 553
XX Db 612 ..... 553
XX
XX Qy 2270 attattcagaatccaacctgaaacacaagttgagctgtgtttccaaggccaattgga 2329
XX ..... 493
XX Db 552 ..... 493
XX
XX Qy 2330 gatattcctagatacctgatagcacgcatgcacaactggaatccagttctcatctt 2389
XX ..... 433
XX Db 492 ..... 433
XX
XX Qy 2390 gaatctcaaaaaatggaccgtctctctatatacgcgttaacctcaaccatgagctcgcgatg 2449
XX ..... 373
XX Db 432 ..... 373
XX
XX Qy 2450 ctgagagatgcccgagttcttactcgcagcagcttcaataaactatggttgacaag 2509
XX ..... 313
XX Db 372 ..... 313
XX
XX Qy 2510 gtgtgtccagttgatagttacaatttggagacaagatccctccagggaagat 2569
XX ..... 253
XX Db 312 ..... 253
XX
XX Qy 2570 gggagggccaagatcctatgttaagattactctccagatagaagtgaataaat 2629
XX ..... 193
XX Db 252 ..... 193
```


THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

us-09-617-099b-2.rqe

| | | | |
|----|----|-----|------------|
| 04 | Db | 474 | RESULT |
| 19 | Qy | 488 | AF19324 |
| 64 | | | LOCUS |
| 79 | Db | 479 | DEFINITION |
| 24 | Qy | 494 | ACCESSION |
| 39 | | | VERSION |
| 04 | | | KEYWORDS |
| 19 | Db | 485 | SOURCE |
| 64 | | | ORGANISM |
| 79 | | | REFERENCE |
| 24 | | | AUTHOR |
| 39 | | | TITLE |
| 04 | | | JOURNALS |
| 19 | | | MEDLINE |
| 64 | | | REFERENCE |
| 79 | | | AUTHOR |
| 24 | | | TITLE |
| 39 | | | JOURNALS |
| 04 | | | FEATURES |
| 19 | | | SOURCE |
| 64 | | | GENE |
| 79 | | | CDS |
| 24 | | | |
| 39 | | | |
| 04 | | | |

10 CAACAACACGCGTTACAAAAACAGAAAGAACACACCCAGGTGG-GGACCCTGTGCCACT 4798
 11 gcatcgtcatgattgtgtcctacagagcccaagctcaggatataccaagctctgttc 4944
 12 GCAATCTGATGATGTTGTCTCTCGGAGCCACAGCTCAGAGGATAAAGACCATCTGTGCTC 4858
 13 tcagagaaga----tcgtacacattgtccctagaagaag 4980
 14 TCAGAGAGAACGCCATCCGTGTCACAGTGTCTCCTTAGCAAGG 4898
 15
 5
 APL199324 5172 bp mRNA ROD 04-JUL-2000
 Rattus norvegicus Rim2-2B (Rim2-2B) mRNA, complete cds.
 APL199324
 APL199324.1 GI:8925861
 Norway rat.
 SM Rattus norvegicus
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 5172)
 Wang, Y., Sugita, S. and Sudhof, T. C.
 The RIM/NIM family of neuronal C2 domain proteins. INTERACTIONS
 WITH RAB3 AND A NEW CLASS OF SRC HOMOLOGUE 3 DOMAIN PROTEINS
 J Biol Chem. 275 (26), 20033-20044 (2000)
 20347919
 2 (bases 1 to 5172)
 Wang, Y. and Sudhof, T. C.
 Direct Submission
 Submitted (27-OCT-1999) Center for Basic Neuroscience, The
 University of Texas Southwestern Medical Center, 6000 Harry Hines
 Blvd., Dallas, TX 75235-9111, USA
 Location/Qualifiers
 1. 5172
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 1. 5172
 /gene="Rim2-2B"
 52..4251
 /gene="Rim2-2B"
 /codon_start=1
 /product="Rim2-2B"
 /protein_id="AAF81646.1"
 /db_xref="GI:8925862"
 /translation="MSAPLAGRGRPAAPPAASOPPEMPDLSTLIEERKLTIVAM
 DGKKEEKEQSVLKRLHOOFVEYKEDVVKMGEESSQOOEOKDATCGICHTKTPAD
 GGGHNSYCOTFCFACGKGRVSLRSKVMYCMCLCKROEILLKSSAMPTNGSSNTPQ
 OPODKALRLBEEAQAQEKAKLHDEOTQOGPGDSSVAVERGRAHGLTRDSINGR
 SGKHOIASDMLBSEARSRFSVSADQNNRYDQSEREYSQVYPSDSMTPEPSYADR
 KQREDEQLEEDPHLWYRDSNRGNHRSKEYIYDDVDSESDREYEQRESEYQAR
 SPTNLARYVAKPOPEEOMRIHAEYSRAHERSHSVSLANAEISRLSLMDPSS
 RORSYSEAAENOPSYEMERTREAOQSSVDPQRTNHSNPPRPSPIPLDEPLR
 ASLIRKHOHLDSASVRKTKRKKMETMLNDLSISDQSSVYRPPPRPHKSKGKKR
 QVLSSESEELASTPEPTSCDVEYETSESVEGKDSQKGRKTSQDAVWSDNTFSER
 OKWVTSGGHSIDEDLEWSEPTKDSGVDTQSTLNEBSHSDDEHPVWQPSKDDR
 LIGRLIKRLNDGVPYRDGAMLGKVVGGKATGESGRCAFTTKKSLADTVGHL
 RGDGDELWNGNLGGAFTFEVYNIILSEKPEQVYLAIVRPTFGDMRPIDSTHGLE
 KSSSESGKMPKRSISVSPSPGMLRDVPOFISGLSIKLMKDFVGHQIITYTIGG
 KLLPSEEGRPKNPNPYKYLEPDRSDKNNRRTKYTKTLEPKNOFTISPVHRSFR
 EMLETTIMDQAVRESESEFAGEITILETLALDDEPWYKLQTHDVSSLVPHRFR
 YMRKQLHSESPTLRKQSRKISDSVSDYDCEDEGAYVSDYRHDRDQSSLTVPYR
 QVMSNHCSPSPSPHRYVIGTRSVSPVPPORNVGALRGTRATGHYNIISMRD
 HVMMDHYSSEDSHFILTPSRHROTSHHHRODGGAAADROPYHRSRSTOYRLL
 ERTTRRSSESEADNLMRSKMTGSRAPSPALRSRSHRTGSSVOTSPSTPYGR
 RROGLPOLPKOTLERGGGKTLRSTVORSTETGLAVENAMNTTQARSSTDSGMSY
 SEEGNLIPGVYALSDQPSDLGGLGQVGTQGLATPAKADLOVGMKRGQLEY
 EITRNGLVYVGGSTLTPAPYKVVLLDNGVGIARKKTVARKLLEPLKQDLSEES
 PGKVLQIILWVDYDRMDKSFMGVAOILIDLELTSMVIGWFKLPPSSLTDPILAP
 LTRKQIISLWSDSTPSYRS"

Page 13

THIS PAGE BLANK (USPTO)

Fri Nov 23 10:08:42 2001

us-09-617-099b-1.rspt

Seq ID NO: 1, Database: SPTREMBL, Page 4
Accession NO: Q9JIS1

141 QELTTSAGMTVSGSNTPOQPOKALRGLESEAPQEKAKLHEOTOFOGPPGDSVPA 200
241 VEGRAHGTLPDQITKNGSVKQIADMPDRKRSPSVSDONRRYQSEREDYSOYV 300
201 VEGRAHGTLPDQITKNGSVKQIADMPDRKRSPSVSDONRRYQSEREDYSOYV 260
301 PSTDYMPSPSDYADRSOREPOFEYEPCHLNYRDSNRRHSHKEYIVDEVEDYSDREY 360
261 PSTDYMPSPSDYADRSOREPOFEYEPCHLNYRDSNRRHSHKEYIVDEVEDYSDREY 320
361 ERORREERYOARYRSDPNLARYPYKPOYEOMRIHAYVSAHRRHSDVSLAAELED 420
321 ERORREERYOARYRSDPNLARYPYKPOYEOMRIHAYVSAHRRHSDVSLAAELED 380
421 SRISLMDRPSRORASYERRAAMENORASYMERTRAOGGSSYPORTSNHSPTRRSP 480
381 SRISLMDRPSRORASYERRAAMENORASYMERTRAOGGSSYPORTSNHSPTRRSP 440
481 IPLDRPMRADSLARKOHNDPSSAVAKTRKEMETMLRNDLSLSDOSEVYRPPRPBK 540
441 IPLDRPMRADSLARKOHNDPSSAVAKTRKEMETMLRNDLSLSDOSEVYRPPRPBK 500
541 SKKGKMRQVSLSSSEELASTPEYSCDVELESESVSEKQKGRKTSSEGVLSDS 600
501 SKKGKMRQVSLSSSEELASTPEYSCDVELESESVSEKQKGRKTSSEGVLSDS 560
601 NTRSEKOKRMATYGGHLEBDELESEPOIKDSGYDTCSTTLNEHSHSDKHPYTWQSK 660
546 -----YSWLEHASM-----SEASPMCLHPYTWQSK 573
661 DGDRLIGRLILKRLKDGVPDPSGAMGLKVVGGKMTESGRLCAFTTKYKKSGLADTVG 720
574 DGDRLIGRLILKRLKDGVPDPSGAMGLKVVGGKMTESGRLCAFTTKYKKSGLADTVG 633
721 HIRPDDEVLENNGRLLQAGTFFEVYNIILASKEPVOVELYVSRPDIPIRIPDSTHAOLE 780
634 HIRPDDEVLENNGRLLQAGTFFEVYNIILASKEPVOVELYVSRPDIPIRIPDSTHAOLE 693
781 SSSSSPESQKMDRPSISVSPSPGMLRDVPOFLSGOLSTIKLMDKXGHOILYVILAKD 840
694 SSSSSPESQKMDRPSISVSPSPGMLRDVPOFLSGOLSTIKLMDKXGHOILYVILAKD 753
841 LPSRBDGRPRNRYVYKLYPDRSDKNKRTKYKTLPEKMNQFIYSPIYHRREREML 900
754 LPSRBDGRPRNRYVYKLYPDRSDKNKRTKYKTLPEKMNQFIYSPIYHRREREML 813
901 ETTIMDOARVREESFELEIILETALIDEPHMYKLOTIDVSSILPLRPPSYLPRRQ 960
814 ETTIMDOARVREESFELEIILETALIDEPHMYKLOTIDVSSILPLRPPSYLPRRQ 873
961 LHGESPTRLQSKRISDSEVSDEYDCEGVVSDYRHNCRDQSSSTLSYPEQVSSNHC 1020
874 LHGESPTRLQSKRISDSEVSDEYDCEGVVSDYRHNCRDQSSSTLSYPEQVSSNHC 933
1021 SPSPGPHRYDVIGTRTSMSPSAPPORANTYEOGHRTATGHYNTISRMDHRYVDDYSS 1080
934 SPSPGPHRYDVIGTRTSMSPSAPPORANTYEOGHRTATGHYNTISRMDHRYVDDYSS 993
1081 DRD-----RDCEADROPYHRSRSTQRLPLERTTSSRSR 1118
994 ERDSHFTLLPSSRHROTSEHHNRDGCDAADROPYHRSRSTQRLPLERTTSSRSR 1053
1119 PDYTNMRSMPSLMTGRSAPPSPALSSHPRTGSVQTSPPSTPCTGCRGROLPOLPPKGT 1178
1054 ADYTNMRSMPSLMTGRSAPPSPALSSHPRTGSVQTSPPSTPCTGCRGROLPOLPPKGT 1113
1179 ER-----SAMDIEBRNOMKLYKQVAGSPRLRLEODYHRSYRSGMDHNG 1224
1114 ERMTEDMDSTRKRNSGANDIEBRNOMKLYKQVAGSPRLRLEODYHRSYRSGMDHNG 1173
1225 ADYVTSKSDSDVSVSAVSRTSSASRSTSYMSVOSERPRGNRIISVYTSKMNQRMG 1284
1174 ADYVTSKSDSDVSVSAVSRTSSASRSTSYMSVOSERPRGNRIISVYTSKMNQRMG 1233

1285 VSGKNTKSTISGDMKSLKNGSOSDITAVGALGTSKRRRSIGAKMVAIVGSRSR 1344
1234 VSGKMAKSTISGDMKSLKNGSOSDITAVGALGTSKRRRSIGAKMVAIVGSRSR 1293
1345 SASOLSTEGGKRLRTVORSTETGLAVENRNMTRASRESTGSKNSYSSEGNLTFP 1404
1294 SASOLSTEGGKRLRTVORSTETGLAVENRNMTRASRESTGSKNSYSSEGNLTFP 1353
1405 GYRLASDSQFDFLDGLGPAVLGROTLAPPMAGDIOGMMDKRGQLEVEIIRANGLYVK 1464
1354 GYRLASDSQFDFLDGLGPAVLGROTLAPPMAGDIOGMMDKRGQLEVEIIRANGLYVK 1413
1465 PGSKTLPAPYVYKYLNDNGVCIARKTKVAKTLEPLYYOQLSEESPOGRVLOIYVGD 1524
1414 PGSKTLPAPYVYKYLNDNGVCIARKTKVAKTLEPLYYOQLSEESPOGRVLOIYVGD 1473
1525 YGRMDHKSFGVQAQIILDELELSNMVQWFKLPSSLYDPTSAPLTRASOSLESSTG 1584
1474 YGRMDHKSFGVQAQIILDELELSNMVQWFKLPSSLYDPTSAPLTRASOSLESSTG 1533
1585 PYSYRS 1590
1534 PYSYRS 1539
RESULT 4
Q9JIS1 PRELIMINARY; PRT; 1555 AA.
AC Q9JIS1;
DT 01-OCT-2000 (TREMREL 15, Created)
DT 01-OCT-2000 (TREMREL 15, Last sequence update)
DT 01-MAR-2001 (TREMREL 16, Last annotation update)
DE R1M2.
GN R1M2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-20347919; PubMed-10748113;
RX Wang Y, Sugita S, Sudhof T.C.;
RT "The R1M/N1M Family of Neuronal C2 Domain Proteins. Interactions with
Rab3 and a new class of Src homology 3 domain proteins.";
RL J. Biol. Chem. 275:20033-20044(2000).
DR EMBL; AF199322; AAF81644.1;
DR InterPro; IPR000008;
DR InterPro; IPR000306;
DR InterPro; IPR000345;
DR InterPro; IPR001478;
DR Pfam; PF00168; C2; 2;
DR Pfam; PF00595; PD2; 1;
DR PROSITE; PS50004; C2 DOMAIN 2; 2;
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1;
DR SMART; SM00239; C2; 1;
SQ SEQUENCE 1555 AA; 175912 MW; D76967BB36D9704E CRC64;
Query Match 89.6%; Score 7420.5; DB 11; Length 1555;
Best Local Similarity 88.4%; Pred. No. 0;
Matches 1451; Conservative 21; Mismatches 31; Indels 139; Gaps 6;
QY 1 MSAPLGPGRAPAPTAASOPPOPEMDLSHTEERKTIILAVMDRQKKEEKQSVLYKI 60
Db 1 MSAPLGPGRAPAPTAASOPPOPEMDLSHTEERKTIILAVMDRQKKEEKQSVLYK 59
QY 61 KEERKAOPTQWPFESGITELVNNVLQPOQKOPNEKEDPTKLHOOFEMYKBOVKMGESQ 120
Db 60 -----KLHOOFEMYKBOVKMGESQ 80
QY 121 QOORXGAPPTGCTGHTKRFADGCGHNGCYCOTKFCARGGGRVSLSNKVMVVCNCLAKO 180
Db 121 QOORXGAPPTGCTGHTKRFADGCGHNGCYCOTKFCARGGGRVSLSNKVMVVCNCLAKO 171

THIS PAGE BLANK (USPTO)

```

Db 81 000BOKDAFTGICHTKTFADGCGHNCYCOTFCACRGRVLSRNKVMWVNLCRQ 140
Qy 181 0EILITSGAMFYNGSMTLQDPDQKVPRLGNEEAPQEKALHEDPOPOGAGDLSVA 240
Db 141 0EILITSGAMFYNGSMTLQDPDQKALGRLSEEAPOEKALHEDTOGAGPDGSSVA 200
Qy 241 VEGRAHGLTFKODITKNGSGVGHQIASDMPDRKSPSPSDQNRREYQSEEREDYSQY 300
Db 201 VEGRAHGLTFKODITKNGSGVGHQIASDMPDRKSPSPSDQNRREYQSEEREDYSQY 260
Qy 301 PSDGTMRSPSDYADRSGREPQFEERPHLNYRDSNRGRHRSKEYIVDEDEVESRDEY 360
Db 261 PSDGTMRSPSDYADRSGREPQFEERPHLNYRDSNRGRHRSKEYIVDEDEVESRDEY 320
Qy 361 ERQREBEYQARYSDPRLARYPVKQPYEEDMRTHAEYSRAHREHSDVSLAMELED 420
Db 321 ERQREBEYQARYSDPRLARYPVKQPYEEDMRTHAEYSRAHREHSDVSLAMELED 380
Qy 421 SRILHMDRPSRQSRVSEERRAAMENORSYSMEPTREAGOGSSYPORTNHSPTPRSP 480
Db 381 SRILHMDRPSRQSRVSEERRAAMENORSYSMEPTREAGOGSSYPORTNHSPTPRSP 440
Qy 481 IPLDRPMRADSLRKQHHLDPSAVERKTKREKMETMLRNDLSLSDQSESVRPPRPHK 540
Db 441 IPLDRPMRADSLRKQHHLDPSAVERKTKREKMETMLRNDLSLSDQSESVRPPRPHK 500
Qy 541 SKKGKARQVLSLSEELASTPEYTCSDVLESESVSEKQSGKRTKSQGVLSDS 600
Db 501 SKKGKARQVLSLSEELASTPEYTCSDVLESESVSEKQSGKRTKSQGVLSDS 560
Qy 601 NTRBERQKKRMYGSHSLDEDLWSEPOIKDQGVDTCSSTTLNEESHSDKHHPVWQPSK 660
Db 561 NTRBERQKKRMYGSHSLDEDLWSEPOIKDQGVDTCSSTTLNEESHSDKHHPVWQPSK 620
Qy 661 DGDRLIGRILNKLKDKSVPRDSCAMGLKAVGKMTESGRICAFITTKYKKSGLADTVG 720
Db 574 DGDRLIGRILNKLKDKSVPRDSCAMGLKAVGKMTESGRICAFITTKYKKSGLADTVG 633
Qy 721 HLRPDEVLENGRLQCATFEVYNIILIESKPEPVELVNSRPIDIPRIPOSTHAQLE 780
Db 634 HLRPDEVLENGRLQCATFEVYNIILIESKPEPVELVNSRPIDIPRIPOSTHAQLE 693
Qy 781 SSSSEFESQKDRPSISVTPSPSGMLRDVPOFLSGOLS-----IKLWF 824
Db 694 SSSSEFESQKDRPSISVTPSPSGMLRDVPOFLSGOLS-----IKLWF 753
Qy 825 DKVGHQILVITIGANDLPREDGRPRNPYKTYFLPDRSDKNRRTKTVKKTLEPKWQOT 884
Db 754 DKVGHQILVITIGANDLPREDGRPRNPYKTYFLPDRSDKNRRTKTVKKTLEPKWQOT 813
Qy 885 FYISVHRRERFERMLETLMDOARVEESEFLGELLILETALDDEPHWYKLOTHDY 944
Db 814 FYISVHRRERFERMLETLMDOARVEESEFLGELLILETALDDEPHWYKLOTHDY 873
Qy 945 SSLPRLPSPFYPRRQLOHGESPTRLQORSKRISDSEVSDVDCEDGVVNSDYRHHNGDLQ 1004
Db 874 SSLPRLPSPFYPRRQLOHGESPTRLQORSKRISDSEVSDVDCEDGVVNSDYRHHNGDLQ 933
Qy 1005 SSTLSVPRQVMSNHCSPSGSPHPRVDYIGTRSWSPSAPPQRNVNVEGHRGTRATGHYNT 1064
Db 934 SSTLSVPRQVMSNHCSPSGSPHPRVDYIGTRSWSPSAPPQRNVNVEGHRGTRATGHYNT 993
Qy 1065 ISRMDRHVRMDHYSDDRD-----RCEADADRPYHRSRSTEQ 1102
Db 994 ISRMDRHVRMDHYSDDRD-----RCEADADRPYHRSRSTEQ 1053
Qy 1103 RPLERRTTSSSEPPDNLMSMPSLMTGRSAPPALSRSHPTGSGVOTSSSTPGT 1162
Db 1054 RPLERRTTSSSEPPDNLMSMPSLMTGRSAPPALSRSHPTGSGVOTSSSTPGT 1113
Qy 1163 GRGRQOLPOLPPKGTLE-----SAMDIEERNRQMLNKYKOVAGSDPRLE 1208
Db 1114 GRGRQOLPOLPPKGTLE-----SAMDIEERNRQMLNKYKOVAGSDPRLE 1173

```

```

Qy 1209 QDYHAKYRSQMDPRGADTVSTKSSDSQSDVSAVSRTSSASRFSSTSYMVSQSERPRGN 1268
Db 1174 QDYHAKYRSQMDPRGADTVSTKSSDSQSDVSAVSRTSSASRFSSTSYMVSQSERPRGN 1233
Qy 1269 RKISVFTSKMONROMGVSGKULTKSTISISGDMCSLEKNDGSDSDRAVGLGTSGRKRSS 1328
Db 1234 RKISVFTSKMONROMGVSGKULTKSTISISGDMCSLEKNDGSDSDRAVGLGTSGRKRSS 1293
Qy 1329 IGAKVAVIYGLSRKSRSSASQSLQTEGGKKLRSYVORSTETGLAVEMRMWMTQASREST 1388
Db 1294 IGAKVAVIYGLSRKSRSSASQSLQTEGGKKLRSYVORSTETGLAVEMRMWMTQASREST 1353
Qy 1389 DGSMSYSSSEGNLIFPGVRLASDSQSFDELDELGAOLVGRQTLATPANGDITGVMMDK 1448
Db 1354 DGSMSYSSSEGNLIFPGVRLASDSQSFDELDELGAOLVGRQTLATPANGDITGVMMDK 1413
Qy 1449 GQLEVEIIRANGLVVPGSKTLPADYKYLIDNGVCIAKKTKVARKTLEPIYQOLLSF 1508
Db 1414 GQLEVEIIRANGLVVPGSKTLPADYKYLIDNGVCIAKKTKVARKTLEPIYQOLLSF 1473
Qy 1509 EESPOGRVLOIIVWGDYGRMDKSPMGVAQIILDELSNMYIGWFKLPPSSLYDPTSA 1568
Db 1474 EESPOGRVLOIIVWGDYGRMDKSPMGVAQIILDELSNMYIGWFKLPPSSLYDPTSA 1533
Qy 1569 PLTRASQSSLESSTGSPYSRS 1590
Db 1534 PLTRASQSSLESSTGSPYSRS 1555

```

```

RESULT 5
09JXSO PRELIMINARY: PRT: 1399 AA.
ID 09JXSO
AC 09JXSO
DT 01-OCT-2000 (TREMBLrel.15, Created)
DT 01-OCT-2000 (TREMBLrel.15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel.16, Last annotation update)
DE RIM2-XB.
GN RIM2-XB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE-203479.9; PubMed-10748113;
RA Wang Y., Sugita S., Sudhof T.C.;
RT "The RIM/NIM Family of Neuronal C2 Domain Proteins: Interactions with
RL J. Biol. Chem. 275:20033-20044(2000).
DR EMBL: AF199324; AAF01646.1;
DR InterPro: IPR000008;
DR InterPro: IPR000306;
DR InterPro: IPR000345;
DR InterPro: IPR001478;
DR Pfam: PF00595; PDZ; 1.
DR PROSITE: PS50004; C2 DOMAIN_2;
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
DR SMART: SM00239; C2; 1.
SQ SEQUENCE 1399 AA; 158945 MW; 3P1550E3D52D237B CRC64;

```

```

Query Match 82.8%; Score 6856.5; DB 11; Length 1399;
Best Local Similarity 83.0%; Pred. No. 0;
Matches 138; Conservative 17; Mismatches 22; Indels 235; Gaps 3;
Qy 1 MSAPLGRGRPATPAASQPPQPEMPDLSHLEEKIILAVNDROKKEEKSQSVLKI 60
Db 1 MSAPLGRGRPATPAASQPPQPEMPDLSHLEEKIILAVNDROKKEEKSQSVLKI 59
Qy 61 KEEHKAQPTQWPEFSGITELVNNVLOPQOKRQNEKEPQTKLHDFEAYKEQYKQGESQ 120
Db 61 KEEHKAQPTQWPEFSGITELVNNVLOPQOKRQNEKEPQTKLHDFEAYKEQYKQGESQ 119

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

0M protein - protein search, using sw model

Run on: November 21, 2001, 16:05:34 ; Search time 43.1 seconds
(without alignments)
4880.858 Million cell updates/sec

Title: US-09-617-099b-1
Perfect score: 8285
Sequence: 1 MSAPLPGRCRAPAPTPAASQP.....TRRASQSSLESSTGSPYSRS 1590

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SPREMBL_16:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_protent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-----------|--------------------|
| 1 | 7930 | 95.7 | 1530 | 11 09EQZ7 | 09eqz7 mus musculu |
| 2 | 7459.5 | 90.0 | 1525 | 11 09JIR5 | 09jir5 rattus norv |
| 3 | 7438.5 | 89.8 | 1539 | 11 09JIR6 | 09jir6 rattus norv |
| 4 | 7420.5 | 89.6 | 1535 | 11 09JIR1 | 09jir1 rattus norv |
| 5 | 6856.5 | 82.8 | 1399 | 11 09JIR0 | 09jir0 rattus norv |
| 6 | 6543 | 79.0 | 1330 | 11 09JIR7 | 09jir7 rattus norv |
| 7 | 6522 | 78.7 | 1352 | 11 09JIR6 | 09jir6 rattus norv |
| 8 | 6504 | 78.5 | 1368 | 11 09JIR9 | 09jir9 rattus norv |
| 9 | 6212 | 75.0 | 1292 | 11 09JIR8 | 09jir8 rattus norv |
| 10 | 5638 | 68.1 | 1188 | 4 09UQ26 | 09uq26 homo sapien |
| 11 | 4513.5 | 54.5 | 1615 | 11 09JIR4 | 09jir4 rattus norv |
| 12 | 4489.5 | 54.2 | 1553 | 11 035168 | 035168 rattus norv |
| 13 | 2936.5 | 35.4 | 1053 | 4 015048 | 015048 homo sapien |
| 14 | 2424.5 | 29.3 | 740 | 4 09HBA5 | 09hba5 homo sapien |
| 15 | 2410.5 | 28.1 | 766 | 4 09HBA6 | 09hba6 homo sapien |
| 16 | 2340.5 | 28.2 | 700 | 4 09HBA4 | 09hba4 homo sapien |
| 17 | 1912 | 23.1 | 369 | 4 03413 | 043413 homo sapien |
| 18 | 1834.5 | 22.1 | 596 | 4 09HBA3 | 09hba3 homo sapien |
| 19 | 1821 | 22.0 | 567 | 4 09HBA2 | 09hba2 homo sapien |

| | | | | | |
|----|--------|------|------|------------|--------------------|
| 20 | 1763.5 | 21.3 | 516 | 4 09HBA1 | 09hba1 homo sapien |
| 21 | 1507.5 | 18.2 | 471 | 4 09HBA3 | 09hba3 homo sapien |
| 22 | 1221.5 | 14.7 | 285 | 11 09JIR2 | 09jir2 rattus norv |
| 23 | 1025.5 | 12.4 | 308 | 4 09JUD0 | 09jud0 homo sapien |
| 24 | 1022 | 12.3 | 307 | 11 09JIR3 | 09jir3 rattus norv |
| 25 | 1021.5 | 12.3 | 308 | 4 092511 | 092511 homo sapien |
| 26 | 1002 | 12.1 | 853 | 5 022366 | 022366 caenorhabdi |
| 27 | 989 | 11.9 | 2464 | 5 09VEE5 | 09vee5 drosophila |
| 28 | 842 | 10.2 | 256 | 4 09H426 | 09h426 homo sapien |
| 29 | 637 | 7.7 | 5120 | 13 09PU36 | 09pu36 gallus gall |
| 30 | 555 | 6.7 | 5085 | 11 09JKS6 | 09jks6 rattus norv |
| 31 | 546.5 | 6.6 | 5038 | 11 09QYX7 | 09qyx7 mus musculu |
| 32 | 500.5 | 6.0 | 743 | 4 043373 | 043373 homo sapien |
| 33 | 475.5 | 5.7 | 1212 | 4 060305 | 060305 homo sapien |
| 34 | 469 | 5.7 | 4880 | 11 09JLRT1 | 09jlr1 rattus norv |
| 35 | 459.5 | 5.5 | 4833 | 11 09QYX6 | 09qyx6 mus musculu |
| 36 | 437 | 5.3 | 1325 | 5 094127 | 094127 caenorhabdi |
| 37 | 371 | 4.5 | 2296 | 4 09UHA8 | 09uha8 homo sapien |
| 38 | 369 | 4.5 | 2752 | 4 09UQ35 | 09uq35 homo sapien |
| 39 | 358 | 4.3 | 547 | 5 021163 | 021163 caenorhabdi |
| 40 | 324.5 | 3.9 | 1262 | 4 09UQ40 | 09uq40 homo sapien |
| 41 | 295 | 3.6 | 3111 | 5 09VH10 | 09vh10 drosophila |
| 42 | 291.5 | 3.5 | 2951 | 5 09W320 | 09w320 drosophila |
| 43 | 290.5 | 3.5 | 5476 | 5 09NJ17 | 09nj17 drosophila |
| 44 | 290.5 | 3.5 | 5533 | 5 09U6C3 | 09u6c3 drosophila |
| 45 | 290.5 | 3.5 | 5533 | 5 09VPL2 | 09vpl2 drosophila |

ALIGNMENTS

| RESULT | 1 | ALIGNMENTS |
|----------------------------|---|--|
| 09EQZ7 | 1 | |
| ID | 09EQZ7 | PRELIMINARY; PRT: 1530 AA. |
| AC | 09EQZ7; | |
| DT | 01-MAR-2001 (TREMBLrel. 16, Created) | |
| DT | 01-MAR-2001 (TREMBLrel. 16, Last sequence update) | |
| DT | 01-MAR-2001 (TREMBLrel. 16, Last annotation update) | |
| DE | RIM2. | |
| GN | RIM2. | |
| OS | Mus musculus (Mouse). | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. | |
| OX | NCBI_TaxID=10090; | |
| RN | [1] | |
| RP | SEQUENCE FROM N.A. | |
| RX | PubMed=11056535; | |
| RA | Ozaki N., Shibasaki T., Kashima Y., Miki T., Takahashi K., Ueno H., | |
| RA | Sunaga Y., Yano H., Matsura Y., Iwanaga T., Takai Y., Saito S.; | |
| RT | "cAMP-GELFI is a direct target of cAMP in regulated exocytosis." | |
| RL | Nat. Cell Biol. 2:805-811(2000). | |
| DR | EMBL; AB021131; BAB18975.1; - | |
| SO | SEQUENCE 1530 AA; 172863 MW; 58CF11BF7152357D CRC64; | |
| Query Match | 95.7%; | Score 7930; DB 11; Length 1530; |
| Best Local Similarity | 96.2%; | Pred. No. 0; |
| Matches 1529; Conservative | 0; | Mismatches 1; Indels 60; Gaps 1; |
| QY | 1 | MSAPLPGRCRAPAPTPAASQPPOPEMDLSHLTEERKIIILAVMDROKKEEKQSYLKI 60 |
| DB | 1 | MSAPLPGRCRAPAPTPAASQPPOPEMDLSHLTEERKIIILAVMDROKKEEKQSYLKI 60 |
| QY | 61 | KEEHAQPTQWPFSGITELVNNVLOPQOKPNEKEBPQRLHOQFEMYKQVKKMGESQ 120 |
| DB | 61 | KEEHAQPTQWPFSGITELVNNVLOPQOKPNEKEBPQRLHOQFEMYKQVKKMGESQ 120 |
| QY | 121 | QOQEKGDAPTCGICHTKTFADGCGHNSYCQKPCARCGRVSIRNKTKMMWCNLCRKQ 180 |
| DB | 121 | QOQEKGDAPTCGICHTKTFADGCGHNSYCQKPCARCGRVSIRNKTKMMWCNLCRKQ 180 |
| QY | 181 | QETLRSGAMFYNSGNTLQOPDQKVPRLRNEEAPQEKKAKLHEDQPOFGAPDLSVPA 240 |
| DB | 181 | QETLRSGAMFYNSGNTLQOPDQKVPRLRNEEAPQEKKAKLHEDQPOFGAPDLSVPA 240 |

```

Db 181 QEILTKSGAMFYNSGNTLQOPDQKVPGRGLRNEAPQOEKKAKLHEQPOFGAGDLSTVPA 240
OY 241 VEGRAHGLTRODITKNGSVKHOIASDMPSPDKRSPSVSRQONRKYEDSEEREDYSQYV 300
Db 241 VEGRAHGLTRODITKNGSVKHOIASDMPSPDKRSPSVSRQONRKYEDSEEREDYSQYV 300
OY 301 PSDGTMPSPSDYADRRSQRPQEFYEEDPGHLNYPDSNRGHHNSKEYIYDDDEVEDSRDEY 360
Db 301 PSDGTMPSPSDYADRRSQRPQEFYEEDPGHLNYPDSNRGHHNSKEYIYDDDEVEDSRDEY 360
OY 361 ERQREEREYQARYSDNRLAYPYKPPQYEEOMRTHAEVSRARHRRHSDVSLAAELED 420
Db 361 ERQREEREYQARYSDNRLAYPYKPPQYEEOMRTHAEVSRARHRRHSDVSLAAELED 420
OY 421 SRSLILMDPPRSRQSVSERAAMENQSYSMERTREAOGOSYQORTSNHSPPTPRSP 480
Db 421 SRSLILMDPPRSRQSVSERAAMENQSYSMERTREAOGOSYQORTSNHSPPTPRSP 480
OY 481 IPLDRPDMRRADSLRKQHLLDPSSAVRKTREKMETMLRNDLSLSDQSESVARPPPRPK 540
Db 481 IPLDRPDMRRADSLRKQHLLDPSSAVRKTREKMETMLRNDLSLSDQSESVARPPPRPK 540
OY 541 SKKGKMKROYSLSSSEELASTPEYTGDDVLESESYSEKGDQSKRKRTSEQVLSDS 600
Db 541 SKKGKMKROYSLSSSEELASTPEYTGDDVLESESYSEKGDQSKRKRTSEQVLSDS 600
OY 601 NTRSERQKKRMYYGSHLEEDLEWSEPOIKDSGYDTCSSTLLNEHSHSDKHPYWPQSK 660
Db 601 NTRSERQKKRMYYGSHLEEDLEWSEPOIKDSGYDTCSSTLLNEHSHSDKHPYWPQSK 660
OY 661 DGRRLIGRIILNKLKDGSPVRODSGAMLGLKVGKMTESGRLCAFITKVKKSLADTVG 720
Db 661 DGRRLIGRIILNKLKDGSPVRODSGAMLGLKVGKMTESGRLCAFITKVKKSLADTVG 720
OY 721 HLRPGGEVLEWNRILQAGAFEEYNYNIILESKPPQVELVYSPRGIDPRIPDSTHAOLE 780
Db 721 HLRPGGEVLEWNRILQAGAFEEYNYNIILESKPPQVELVYSPRGIDPRIPDSTHAOLE 780
OY 781 SSSSESESKMDRPSISVTPSPMSPGMLRDVPQFLSGOLSIKLMFDKVGHOILVITLGAKD 840
Db 781 SSSSESESKMDRPSISVTPSPMSPGMLRDVPQFLSGOLSIKLMFDKVGHOILVITLGAKD 840
OY 841 LPSREGGRPNRPVYKTYFLPDRSDKNNKRTKYVAKTLEPKNNQTFIYSPVHRREFRML 900
Db 841 LPSREGGRPNRPVYKTYFLPDRSDKNNKRTKYVAKTLEPKNNQTFIYSPVHRREFRML 900
OY 901 ETLMDQARVRESESEFLGILIEFTALLDDPEHMYKLOTHDVSSPLPRPSPTLPRQ 960
Db 901 ETLMDQARVRESESEFLGILIEFTALLDDPEHMYKLOTHDVSSPLPRPSPTLPRQ 960
OY 961 LHGESPTRRLQSRKRTSDSEVSDYDCEDGVGVVSDYRHHNGRDLOSSLVPEQVWSSNHC 1020
Db 961 LHGESPTRRLQSRKRTSDSEVSDYDCEDGVGVVSDYRHHNGRDLOSSLVPEQVWSSNHC 1020
OY 1021 SPGSGPHRYDVIGRTKSWSPAPPPQANVQGRGTATGHYNTISMDRRRVNDHYS 1080
Db 1021 SPGSGPHRYDVIGRTKSWSPAPPPQANVQGRGTATGHYNTISMDRRRVNDHYS 1080
OY 1081 DHRDRCQEAADROPYHRSRSTQRPLLERTTTRSSSERPDTNLRMSPLMTGSAVPS 1140
Db 1081 DHRDRCQEAADROPYHRSRSTQRPLLERTTTRSSSERPDTNLRMSPLMTGSAVPS 1140
OY 1141 ALSRSHPTGVSQVTPSSPTPGTGRGRQPLPQLPKGTGLERSAMDIEERNQMLNKYKQY 1200
Db 1141 ALSRSHPTGVSQVTPSSPTPGTGRGRQPLPQLPKGTGLERSAMDIEERNQMLNKYKQY 1200
OY 1201 AGSDRLLEDYSHKYSAGMDPHRGADTVSTKSSDSYDVSAVSRSTSASRFSSTSTMSV 1260
Db 1201 AGSDRLLEDYSHKYSAGMDPHRGADTVSTKSSDSYDVSAVSRSTSASRFSSTSTMSV 1260
OY 1261 QSERPRGNRKISVFTSKMONRQMGVSGKNLTKSTISGDMCSLEKNDGSDOTAVGLGT 1320
Db 1261 QSERPRGNRKISVFTSKMONRQMGVSGKNLTKSTISGDMCSLEKNDGSDOTAVGLGT 1320
OY 1321 QSERPRGNRKISVFTSKMONRQMGVSGKNLTKSTISGDMCSLEKNDGSDOTAVGLGT 1380
Db 1321 QSERPRGNRKISVFTSKMONRQMGVSGKNLTKSTISGDMCSLEKNDGSDOTAVGLGT 1380
OY 1381 ROASRESTDGSMSYSEGNLIFPGVRLASQSFDFLDGSPAOVLGROTLAPAMDI 1440
Db 1381 ROASRESTDGSMSYSEGNLIFPGVRLASQSFDFLDGSPAOVLGROTLAPAMDI 1440
OY 1441 QVGMDDKKGQLEVEILIRAGLVKPGSKTLPAPYKYLNDNGVCIAKKTKVARTLEP 1500
Db 1441 QVGMDDKKGQLEVEILIRAGLVKPGSKTLPAPYKYLNDNGVCIAKKTKVARTLEP 1500
OY 1501 LYQOOLSFEESSPGQRYVQIITWGDYGRMDKSPFNGVQOILDELELSNMVIGFKLPPS 1560
Db 1501 LYQOOLSFEESSPGQRYVQIITWGDYGRMDKSPFNGVQOILDELELSNMVIGFKLPPS 1560
OY 1561 SLVDPTAPLTRRASQSSLESSTGSPYSRS 1590
Db 1501 SLVDPTAPLTRRASQSSLESSTGSPYSRS 1590
OY 1590 SLVDPTAPLTRRASQSSLESSTGSPYSRS 1650
Db 1501 SLVDPTAPLTRRASQSSLESSTGSPYSRS 1650

```

```

OY 1321 SGKKRRSSITAKKVAIVGLSKRSSASQALSQTEGGKKLRSTVQSTETGLAVENRMMT 1380
Db 1261 SGKKRRSSITAKKVAIVGLSKRSSASQALSQTEGGKKLRSTVQSTETGLAVENRMMT 1320
OY 1381 ROASRESTDGSMSYSEGNLIFPGVRLASQSFDFLDGSPAOVLGROTLAPAMDI 1440
Db 1321 ROASRESTDGSMSYSEGNLIFPGVRLASQSFDFLDGSPAOVLGROTLAPAMDI 1380
OY 1441 QVGMDDKKGQLEVEILIRAGLVKPGSKTLPAPYKYLNDNGVCIAKKTKVARTLEP 1500
Db 1381 QVGMDDKKGQLEVEILIRAGLVKPGSKTLPAPYKYLNDNGVCIAKKTKVARTLEP 1440
OY 1501 LYQOOLSFEESSPGQRYVQIITWGDYGRMDKSPFNGVQOILDELELSNMVIGFKLPPS 1560
Db 1441 LYQOOLSFEESSPGQRYVQIITWGDYGRMDKSPFNGVQOILDELELSNMVIGFKLPPS 1500
OY 1561 SLVDPTAPLTRRASQSSLESSTGSPYSRS 1590
Db 1501 SLVDPTAPLTRRASQSSLESSTGSPYSRS 1590

RESULT 2
O9JIR5 PRELIMINARY; PRT: 1525 AA.
AC O9JIR5:
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, last sequence update)
DT 01-MAR-2001 (TEMBLrel. 16, last annotation update)
DE RIM2-5B.
GN RIM2-5B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347919; PubMed=10748113;
RA Wang Y., Sugita S., Sudhof T.C.;
RT "The RIM/NIM Family of Neuronal C2 Domain Proteins. Interactions with
RT Rabs and a new class of Src homology 3 domain proteins.";
RL J. Biol. Chem. 275:20033-20044(2000).
DR EMBL; AF199331; AAF81653.1; -.
DR InterPro; IPR000008; -.
DR InterPro; IPR000306; -.
DR InterPro; IPR000345; -.
DR InterPro; IPR001478; -.
DR Pfam; PF00168; C2; 2.
DR Pfam; PF00595; PD2; 1.
DR PROSITE; PS00004; C2_DOMAIN_2; 2.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR SMART; SM00239; C2; 1.
SQ SEQUENCE 1525 AA; 172435 MW; D4930D62F03CF21 CRC64;

Query Match 90.0%; Score 7459.5; DB 11; Length 1525;
Best Local Similarity 90.1%; Pred. No. 0;
Matches 1452; Conservative 21; Mismatches 30; Indels 109; Gaps 4;

OY 1 MSAPLPGRPAPTPAASQPPQPEMPDLISLTHEERKILAYMDROKKEEKEQSVLKI 60
Db 1 MSAPLPGRPAPTPAASQPPQPEMPDLISLTHEERKILAYMDROKKEEKEQSVLKI 60
OY 61 KEENHAQPTQWPFPGSITLVANNVLQPOOKQRENEKQPQTKLHQEFMYKEQYKMGEEQ 120
Db 60 -----KLHQEFMYKEQYKMGEEQ 80
OY 121 QOQEQKGAAPTQIGCHKTFTFADGCHNCSCYQTKFCARGGRVSLRSNKVVMVVCNLCRQ 180
Db 81 QOQEQKGAAPTQIGCHKTFTFADGCHNCSCYQTKFCARGGRVSLRSNKVVMVVCNLCRQ 140
OY 181 QEILTKSGAMFYNSGNTLQOPDQKVPGRGLRNEAPQOEKKAKLHEQPOFGAGDLSTVPA 240
Db 181 QEILTKSGAMFYNSGNTLQOPDQKVPGRGLRNEAPQOEKKAKLHEQPOFGAGDLSTVPA 240

```

```
Db 141 QELLTSGAMFYNSGNTPPQOPDOKALRGLRSEAPQOEKAKLHEOTQOFQPGSDSVPA 200
Qy 241 VEKRAHGILTRDITJKNKGCVKHOLASDMPDREKRSVPVSARDONRRYEOGEEREDYSQY 300
Db 201 VERGAHGLITRODSTJKNKGCKHQLASDMPDREKRSVPVSARDONRRYEOGEEREDYSQY 260
Qy 301 PSDGTMPSPSDYADRRSOREPOFEPECHLNYRDSNRGRHRSKEYITVDEDEVEDSDEY 360
Db 261 PSDTMRPSPDYADRRSOREPOFEPECHLNYRDSNRGRHRSKEYITVDEDEVEDSDEY 320
Qy 361 ERORREEEQARYRSDPULARYPVKQPYEEOMRIHAENSRARHRRHSDVSLANAELED 420
Db 321 ERORREEEQARYRSDPULARYPVKQPYEEOMRIHAENSRARHRRHSDVSLANAELED 380
Qy 421 SRISLLRDRPORSVSRRRAMENORSYMERTEAOGSSYPOSTSNHSPPTPRSP 480
Db 381 SRISLLRDRPORSVSRRRAMENORSYMERTEAOGSSYPOSTSNHSPPTPRSP 440
Qy 481 IPLDRPMRADSLRKQHLDPSSAVRKTREKEMTMRDNDLSLSDQSESVRPPRPBK 540
Db 441 IPLDRPMRADSLRKQHLDPSSAVRKTREKEMTMRDNDLSLSDQSESVRPPRPBK 500
Qy 541 SKGGKMRQVSLSSSEELASTPEYTSODVLESESVSEKQSGKRTSSEGVLSDS 600
Db 501 SKGGKMRQVSLSSSEELASTPEYTSODVLESESVSEKQSGKRTSSEGVLSDS 560
Qy 601 NTRSRQKRMVYGGSLSEDELEWSEPOJKDQSVDCSSITLNEHSHSDKHPTWQPSK 660
Db 566 -----Y5WLEHASW-----S5EASPM5LHPPTWQPSK 573
Qy 661 DGDRLIGRILLNKLKDGSVPRDQSGAMLGKVVCGKMTESGRICAFITKYKKSGLADTVG 720
Db 574 DGDRLIGRILLNKLKDGSVPRDQSGAMLGKVVCGKMTESGRICAFITKYKKSGLADTVG 633
Qy 721 HLRPDEVLWNGRLQGAFTFEEVYNIIEESKPEOVELVSRPQDIPRIPOSTHAOLE 780
Db 634 HLRPDEVLWNGRLQGAFTFEEVYNIIEESKPEOVELVSRPQDIPRIPOSTHAOLE 693
Qy 781 SSSSFEESQKMRPSTSVSPMSGMLRQVPOFLSGOLSTIKLWFDKQGHOLITVILAKD 840
Db 694 SSSSFEESQKMRPSTSVSPMSGMLRQVPOFLSGOLSTIKLWFDKQGHOLITVILAKD 753
Qy 841 LPSRDGGRPNRYVYKIFLPDRSDDKKRRKTIVYKTLLEPMNTOFTIYSPVHRREFRRL 900
Db 754 LPSRDGGRPNRYVYKIFLPDRSDDKKRRKTIVYKTLLEPMNTOFTIYSPVHRREFRRL 813
Qy 901 EITLWDQARVRESEEFELGELLIELETALLDDEPHMYKLTQHDVSSLPDRPSEYLRQ 960
Db 814 EITLWDQARVRESEEFELGELLIELETALLDDEPHMYKLTQHDVSSLPDRPSEYLRQ 873
Qy 961 LHGESPTRLQSKRISDSVSDYDCEDEGVGVSDYRHNGRDLQSSFLSVPEQYVSSNHC 1020
Db 874 LHGESPTRLQSKRISDSVSDYDCEDEGVGVSDYRHNGRDLQSSFLSVPEQYVSSNHC 933
Qy 1021 SPSSGPHRDVIGRTSRSPSPAPPQORNYEOGHRGTFRATGHYNTISPMDRHRVNDHXS 1080
Db 934 SPSSGPHRDVIGRTSRSPSPAPPQORNYEOGHRGTFRATGHYNTISPMDRHRVNDHXS 993
Qy 1081 DRD-----RDCEAADRQYHRSRSTEDRPLLETTTTRSSER 1118
Db 994 EEDSHFLTLPRSRHROTSEHHNRGDCDEAADRQYHRSRSTEDRPLLETTTTRSSER 1053
Qy 1119 PDTNLMRSPSLMTGSAAPSPALSRSHPTGSGVOTSPSSPTPGTGRGRQOLPQRPKGT 1178
Db 1054 ADTNLMRSPSLMTGSAAPSPALSRSHPTGSGVOTSPSSPTPGTGRGRQOLPQRPKGT 1113
Qy 1179 ERSADIEERNOMKLINKKQYAGSDPRLQDYHSHKYSRGMDDPHRGADYVSTKSSDSVS 1238
Db 1114 ERSADIEERNOMKLINKKQYAGSDPRLQDYHSHKYSRGMDDPHRGADYVSTKSSDSVS 1173
Qy 1239 DVSASVRSSTSSASRFSSTYSVOSERPGRNKISVFTSKMNQMGVSGKNLTKSTISG 1298
Db 1174 DVSASVRSSTSSASRFSSTYSVOSERPGRNKISVFTSKMNQMGVSGKNLTKSTISG 1233
```

```
Qy 1299 DMC5LEKNDGSDFAVAGALGTSGKKRRSSIGAKMAIVGLSRKSRASQLSQTEGGKK 1358
Db 1234 DMC5LEKNDGSDFTAVAGALGTSGKKRRSSIGAKMAIVGLSRKSRASQLSQTEGGKK 1293
Qy 1359 LRSTVORSTETGLAVEMRMNMTROASRESTDGSNMNSYSEGNLFFPGVRLASDSQSDFL 1418
Db 1294 LRSTVORSTETGLAVEMRMNMTROASRESTDGSNMNSYSEGNLFFPGVRLASDSQSDFL 1353
Qy 1419 DLGLPQVLGRQTLATPAGDIOVGMADKKQLEVEILIRAGLVYKSGSKLPPAPYKYV 1478
Db 1354 DLGLPQVLGRQTLATPAGDIOVGMADKKQLEVEILIRAGLVYKSGSKLPPAPYKYV 1413
Qy 1479 LLDNGVCIKKKTIVARKTLEPLXOOLLSPEESQGRVLOIIVYGDYGRMDHKSGFMGYAO 1538
Db 1414 LLDNGVCIKKKTIVARKTLEPLXOOLLSPEESQGRVLOIIVYGDYGRMDHKSGFMGYAO 1473
Qy 1539 ILDELELSNMVIGWFKLPPSSLYDPTSAPLTRRASQSSLESSTGSPSYRS 1590
Db 1474 ILDELELSNMVIGWFKLPPSSLYDPTSAPLTRRASQSSLESSTGSPSYRS 1525

RESULT 3
QyIR6 PRELIMINARY; PRT; 1539 AA.
ID Q9JIR6
AC Q9JIR6;
DT 01-OCT-2000 (TRENBLREL, 15, Created)
DT 01-OCT-2000 (TRENBLREL, 15, Last sequence update)
DT 01-MAR-2001 (TRENBLREL, 16, Last annotation update)
DE RIM-5A.
GN RIM-5A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20347919; PubMed-10748113;
RA Wang Y., Sugita S., Sudhof T.C.;
RT "The RIM/NIM Family of Neuronal C2 Domain Proteins. Interactions with
RT Rab3 and a new class of Src homology 3 domain proteins.";
RL J. Biol. Chem. 275:20033-20044(2000).
DR EMBL: AF199330; AAF81652.1; -.
DR InterPro: IPR000008; -.
DR InterPro: IPR000306; -.
DR InterPro: IPR000345; -.
DR InterPro: IPR001478; -.
DR Pfam: PF00168; C2; 2.
DR Pfam: PF00595; PDZ; 1.
DR PROSITE: PS00004; C2 DOMAIN_2; 2.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
DR SMART: SM00239; C2; 1.
SQ SEQUENCE 1539 AA; 174071 MW; 12882D2F4E0AE8C CRC64;

Query Match 89.8%; Score 7438.5; DB 11; Length 1539;
Best Local Similarity 89.2%; Pred. No. 0;
Matches 1451; Conservative 21; Mismatches 31; Indels 123; Gaps 5;
```

```
Db 141 OEILTSGAMFYNSGNTPOQDQKALGLRSFEAPQEKAKLHEQTOFOGPPGSDSVA 200
QY 241 VEGRAHAGLIRQOTIKNGSGVKHOLASDMPDRKRSVSDQNRNRYEQSEREDYSQYV 300
Db 201 VEGRAHAGLIRQOSIKNGSGKHOLASDMPDRKRSVSDQNRNRYEQSEREDYSQYV 260
QY 301 PSDGTPRSPSDVADRKSOREPOFYEEFGHLNVRDNRGRHRSKEYIVDDDVSRDEY 360
Db 261 PSDGTPRSPSDVADRKSOREPOFYEEFGHLNVRDNRGRHRSKEYIVDDDVSRDEY 320
QY 361 ERORREBEVOARYRSDBNLARYPVKPOPYEQOMRIHAEVSAARHERHSDVSLANAELD 420
Db 321 ERORREBEVOARYRSDBNLARYPVKPOPYEQOMRIHAEVSAARHERHSDVSLANAELD 380
QY 421 SRTSLIRMDPSRORSYSEERAAMENORSYMERETREAOGOSYPOKRTSNHSPPTPRSP 480
Db 381 SRTSLIRMDPSRORSYSEERAAMENORSYMERETREAOGOSYPOKRTSNHSPPTPRSP 440
QY 481 IPLDRPDMRADSLRKOHHLDPSSAVRKTREKMETMLRNDLSLSDQSESVPRPPRPHK 540
Db 441 IPLDRPDMRADSLRKOHHLDPSSAVRKTREKMETMLRNDLSLSDQSESVPRPPRPHK 500
QY 541 SKKGKMKROVSLSSSEBELASTPEYTSODVLESESVSEKGDOKGKRKTSQGVLSDS 600
Db 501 SKKGKMKROVSLSSSEBELASTPEYTSODVLESESVSEKGDOME----- 545
QY 601 NTFSEKOKKMYTGGHLEEDLEWSEPOIKDSGYDTCSTTLNHEHSHDKHPYWOBSK 660
Db 546 -----XSWLEHSAWH-----SSASPMSLHPYWOBSK 573
QY 661 DGGRLIGRILNLRKLDGVPYRDSGAMGLKVYGGKMTESGRICAFITKVKGSTADYV 720
Db 574 DGGRLIGRILNLRKLDGVPYRDSGAMGLKVYGGKMTESGRICAFITKVKGSTADYV 633
QY 721 HLRPGDEVLWNGRLLQGAFFEEYNNILLESKPEPOVELVSRPIGDIPLRIPDSTHAOLE 780
Db 634 HLRPGDEVLWNGRLLQGAFFEEYNNILLESKPEPOVELVSRPIGDMRIPDSTHAOLE 693
QY 781 SSSSSPESOKMDRPSISVTPSPSGMLRDVPOFLSGOLSTLMPKVGHOHLVITLIGAD 840
Db 694 SSSSSPESOKMDRPSISVTPSPSGMLRDVPOFLSGOLSTLMPKVGHOHLVITLIGAD 753
QY 841 LPSREDGRPNRPYKITYFLPDRSDKNRKRTKYKTKLEPPKNNQFTIYSPVHREPEREML 900
Db 754 LPSREDGRPNRPYKITYFLPDRSDKNRKRTKYKTKLEPPKNNQFTIYSPVHREPEREML 813
QY 901 ETTLMQOARKRESESEFLGELLLELFTALLDDEPHWYKLQTHDVSSLPPLRPSYLPFRQ 960
Db 814 ETTLMQOARKRESESEFLGELLLELFTALLDDEPHWYKLQTHDVSSLPPLRPSYLPFRQ 873
QY 961 LHGESPTRLORSKRISDSSEVDYDCEDGVGVSDYRHNGRDLQSTLSVPEQVWSSNHC 1020
Db 874 LHGESPTRLORSKRISDSSEVDYDCEDGVGVSDYRHNGRDLQSTLSVPEQVWSSNHC 933
QY 1021 SPGSGPHRDVIGITRSMSPSAPPPORNVDEGHRGTRATGHTNTISRMDRHRVMDHYS 1080
Db 934 SPGSGPHRDVIGITRSMSPSAPPPORNVDEGHRGTRATGHTNTISRMDRHRVMDHYS 993
QY 1081 DRD-----RDCEAADROPYHRSRSTEOPLLERTTTRSSSER 1118
Db 994 ERDSHFLLTPRSRHRQTESEHHNRDRDCEAADROPYHRSRSTEOPLLERTTTRSSSER 1053
QY 1119 PPTNLMRSMPSLMTGSRAPSPALSRSHPTGSAVOTSPSSPTGGRGRGROLPOLPPKCTL 1178
Db 1054 ADTNLMKSMPSLMTGSRAPSPALSRSHPTGSAVOTSPSSPTGGRGRGROLPOLPPKCTL 1113
QY 1179 ER-----SAMDIEERNROMKLNKYQOVAGSDPRLLEODYISKXRSAGDPRHG 1224
Db 1114 ERMITEEDMDSTRKRNAGANDIEERNROMKLNKYQOVAGSDPRLLEODYISKXRSAGDPRHG 1173
QY 1225 ADTVSTKSSDSVDYSAVSRSTSSAFRSTSYWVOSERPGRNKKISVFTSKMOROMG 1284
Db 1174 ADTVSTKSSDSVDYSAVSRSTSSAFRSTSYWVOSERPGRNKKISVFTSKMOROMG 1233
```

```
QY 1285 VSGKNLTKSTISGDMCSLEKNDGSDPTAVGALGTSCKKRRRSTGAKMVAIVGLSRKS 1344
Db 1234 VSGKSMKSTISGDMCSLEKNDGSDPTAVGALGTSCKKRRRSTGAKMVAIVGLSRKS 1293
QY 1345 SASOLQTEEGGKKLRSTVQSTETGLAVERNMNTROASREESTGSKNSYSSEGNLFFP 1404
Db 1294 SASOLQTEEGGKKLRSTVQSTETGLAVERNMNTROASREESTGSKNSYSSEGNLFFP 1353
QY 1405 GVRLASDSQSFDELIDGLGPAOLVGRQTLATPAMGDIOVGMDKRGQLEVEITRAGLYVK 1464
Db 1354 GVRLASDSQSFDELIDGLGPAOLVGRQTLATPAMGDIOVGMDKRGQLEVEITRAGLYVK 1413
QY 1465 PGSKTLPPAPYKYYLLDNGVCIAKKTKRVAKKTLEPLYOOLLSPEESPQGRVLOIIVMGD 1524
Db 1414 PGSKTLPPAPYKYYLLDNGVCIAKKTKRVAKKTLEPLYOOLLSPEESPQGRVLOIIVMGD 1473
QY 1525 YGRMDHKSFMGVAAQIILDELELSNMVIGWFELFPPSSLVDP TSAPLTRRASQSSLESSTG 1584
Db 1474 YGRMDHKSFMGVAAQIILDELELSNMVIGWFELFPPSSLVDP TSAPLTRRASQSSLESSTG 1533
QY 1585 PSTSRS 1590
Db 1534 PSTSRS 1539

RESULT 4
Q9JIS1 PRELIMINARY; PRT; 1555 AA.
ID Q9JIS1;
AC Q9JIS1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE RIM2.
GN RIM2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347919; PubMed=10748113;
RA Wang Y., Sugita S., Sudhof T.C.;
RT "The RIM/NIM Family of Neuronal C2 Domain Proteins. Interactions with
RT Rab3 and a new class of Src homology 3 domain proteins.";
RL J. Biol. Chem. 275:20033-20044(2000).
DR EMBL; AF199322; AAF81644.1; -.
DR InterPro; IPR000008; -.
DR InterPro; IPR000306; -.
DR InterPro; IPR000345; -.
DR InterPro; IPR001478; -.
DR Pfam; PF00168; C2; 2.
DR Pfam; PF00595; PDZ; 1.
DR PROSITE; PS50004; C2_DOMAIN_2; 2.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR SMART; SM00239; C2; 1.
SQ SEQUENCE 1555 AA; 175912 MW; D76967BB36D9704E CRC64;
```

Query Match 89.6%; Score 7420.5; DB 11; Length 1555;
Best local similarity 88.4%; Pred. No. 0;
Matches 1451; Conservative 21; Mismatches 31; Indels 139; Gaps 6;

```
QY 1 MSAPLGRGRPAPTPAASQPPPEMPDLHLTEERKIIIAVMDROKKEEKEQSVLKI 60
Db 1 MSAPLGRGRPAPTPAASQPPPEMPDLHLTEERKIIIAVMDROKKEEKEQSVLKI 59
QY 61 KEENKAAQPTQWPFPSGTTELNVNVLQPOQKPNKEPQTKLHQOFEMKKEQVKKGEESO 120
Db 60 -----KLHQOFEMKKEQVKKGEESO 80
QY 121 QOQEOKGAFTCGICHTKTRADGCHNCYQOTFCACRGGRVLSRNVAVNCIMCRQ 180
Db 121 QOQEOKGAFTCGICHTKTRADGCHNCYQOTFCACRGGRVLSRNVAVNCIMCRQ 180
```

```

Db 81 QOQOBKGDAPTGCIGCHTKFADGCGHNCSCQTCFARCGGRVLSRKNVWVYCNLCRKQ 140
QY 181 QELLTKSGAMFYNSGNTLQOPDOKVPRGLRMEBAPOEKKAKLHEOPQOPGAGDLSVPA 240
Db 141 QELLTKSGAMFYNSGNTLQOPDOKALKRLSEEAPOEKKAKLHEOTOPOGPGDSSVPA 200
QY 241 VEKGRAGILTRQDTIKNGSGVHQIASDMPSDRKSPSVSDQNRREYEDSEEREDYSQY 300
Db 201 VERKRAHOLTQDSTIKNGSGMKHQIASDMPSDRKSPSVSDQNRREYEDSEEREDYSQY 260
QY 301 PSDGTMPSPDYADRSRQREPQYEEPGHLYRDSNRGRHRSKEYIVDEDEVEDSEDEX 360
Db 261 PSDGTMPSPDYADRSRQREPQYEEPDHLYRDSNRGRHRSKEYIVDEDEVEDSEDEX 320
QY 361 EROREBEYQARYSDPRLARYPVKPOPYEEQMRTHAEVSRAHRRHSDVSLANMELED 420
Db 321 EROREBEYQARYSDPRLARYPVKPOPYEEQMRTHAEVSRAHRRHSDVSLANMELED 380
QY 421 SRIILRMDRPSRQSVSEERRAAMENORSYMERTREAGOSSYPORTSNHSPPTPRSP 480
Db 381 SRIILRMDRPSRQSVSEERRAAMENORSYMERTREAGOSSYPORTSNHSPPTPRSP 440
QY 481 IPLDRPDMRADSLRKQHHLDPSAVRKTREKMETMLRNDLSLSDQSESVRPPRPK 540
Db 441 IPLDRPDMRADSLRKQHHLDPSAVRKTREKMETMLRNDLSLSDQSESVRPPRPK 500
QY 541 SKGGMKQVSLSSSEELASTPEYTCSDVLESESVSEKDGSKGKRTSQGVLSDS 600
Db 501 SKGGMKQVSLSSSEELASTPEYTCSDVLESESVSEKDGME----- 545
QY 601 NTRSERQKRMYYGSHSLSEEDLEMSSEPOLKDGSDVTCSTTLNEHSHSDKHPTWQPSK 660
Db 546 -----YSMLHSAHWH-----SSEASPMASHVYTQPSK 573
QY 661 DGDRLIGRILLNKLKDKSVPRDSCAMGLKLVYGGKMTESGRICAFITTKYKKSGLADTVG 720
Db 574 DGDRLIGRILLNKLKDKSVPRDSCAMGLKLVYGGKMTESGRICAFITTKYKKSGLADTVG 633
QY 721 HLRRGDEVLENGRLLQCATPEEYVNIILLESKPEPOVELYVSNRPIDIRIPDSTHAOLE 780
Db 634 HLRRGDEVLENGRLLQCATPEEYVNIILLESKPEPOVELYVSNRPIDIPDSTHAOLE 693
QY 781 SSSSFESQKMDRPSIYSTSPKPMGLRDVPOFLSGOLS-----IKLWF 824
Db 694 SSSSFESQKMDRPSIYSTSPKPMGLRDVPOFLSGOLS-----IKLWF 753
QY 825 DKVGHQILVTLLGAKDLPSREDGRPRNPVYKIFPLPDRSDKNKRRTKYVKTLEPKWQJ 884
Db 754 DKVGHQILVTLLGAKDLPSREDGRPRNPVYKIFPLPDRSDKNKRRTKYVKTLEPKWQJ 813
QY 885 FIYSVPHRRERERLETTIMDOARVRESESEFLGILTELEFALLDDPRHMKLOTHDV 944
Db 814 FIYSVPHRRERERLETTIMDOARVRESESEFLGILTELEFALLDDPRHMKLOTHDV 873
QY 945 SSLPLPRSPYLPRLRQLGESPTRRLOSKRISLSEVSDYDCEDGCVVSDYVHHNGRDLQ 1004
Db 874 SSLPLPRSPYLPRLRQLGESPTRRLOSKRISLSEVSDYDCEDGCVVSDYVHHNGRDLQ 933
QY 1005 SSTLSVPEQVWSSNHCSPGSHRVADVIGRTKSWSPSAPPPORANVOGHGRTATGHTNT 1064
Db 934 SSTLSVPEQVWSSNHCSPGSHRVADVIGRTKSWSPSAPPPORANVOGLGRTATGHTNT 993
QY 1065 ISRMDRHRVMDHYSSDD-----RCEAADRPOYHRSRTEQ 1102
Db 994 ISRMDRHRVMDHYSSDDSHFLTLPRSRHROTSEHNRDGRCEADROPYHRSRTEQ 1053
QY 1103 RPLLERTTSSRSSEPTNLKMSMPSLMTGSAAPSPALSRSHPRTGSVOTSPSSPTGT 1162
Db 1054 RPLLERTTSSRSSEPTNLKMSMPSLMTGSAAPSPALSRSHPRTGSVOTSPSSPTGT 1113
QY 1163 GRGRQOLPOLPPKGLTER-----SANDIEBRNROMKLNKYQOVAGSDPRL 1208
Db 1114 GRGRQOLPOLPPKGLTEREDMDSTRKNSGAMDIEBRNROMKLNKYQOVAGSDPRL 1173

```

```

QY 1209 ODYHSKYSRGMDFHRGADPTVSTRKSSDSDVDVSAVSRSTSSAPFSSSTYSVOSERPRGN 1268
Db 1174 ODYHSKYSRGMDFHRGADPTVSTRKSSDSDVDVSAVSRSTSSAPFSSSTYSVOSERPRGN 1233
QY 1269 KRISVETSKMONRQMGVSGKNLTKSTISIGDMCSLEKNDGSDTAVGALGTSGKKRRSS 1328
Db 1224 KRISVETSKMONRQMGVSGKNLTKSTISIGDMCSLEKNDGSDTAVGALGTSGKKRRSS 1293
QY 1329 IGAKNVAIVGLSKRSASQLSQTEBGKKLSTVORSTETGLAVEMRNMTROASREST 1388
Db 1294 IGAKNVAIVGLSKRSASQLSQTEBGKKLSTVORSTETGLAVEMRNMTROASREST 1353
QY 1389 DGSNMSYSEGNLTFPGVGLASDSQFSDLGIPGPOIVGROFLAPAMGDIOVGMMDK 1448
Db 1354 DGSNMSYSEGNLTFPGVGLASDSQFSDLGIPGPOIVGROFLAPAMGDIOVGMMDK 1413
QY 1449 GQLEVEIIRARGLVYKPGSKTLPAPYVKKYLLDNGVCIKKTKVARKTLEPLYOQLLSF 1508
Db 1414 GQLEVEIIRARGLVYKPGSKTLPAPYVKKYLLDNGVCIKKTKVARKTLEPLYOQLLSF 1473
QY 1509 EESPOGRVLOIIVMGDYGMDHKSFGVQAQILDELELSNMVIGWEKLEPPSSLVDPSTA 1568
Db 1474 EESPOGRVLOIIVMGDYGMDHKSFGVQAQILDELELSNMVIGWEKLEPPSSLVDPSTA 1533
QY 1569 PLTRRASOSLESSTGSPYSRS 1590
Db 1534 PLTRRASOSLESSTGSPYSRS 1555

RESULT 5
O9JISO PRELIMINARY: PRT: 1399 AA.
AC O9JISO:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE RIM2-2B.
GN RIM2-2B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=20347919; Pubmed=10748113;
RA Wang Y., Sugita S., Sudhof T.C.;
RT "The RIM/NIM Family of Neuronal C2 Domain Proteins: Interactions with
RT Rab3 and a new class of Src homology 3 domain proteins.";
RL J. Biol. Chem. 275:20033-20044(2000).
DR EMBL; AF199324; AAF81646.1; -.
DR InterPro; IPR000008; -.
DR InterPro; IPR000306; -.
DR InterPro; IPR000345; -.
DR InterPro; IPR001478; -.
DR Pfam; PF00168; C2; 2.
DR Pfam; PF00595; PDZ; 1.
DR PROSITE; PS50004; C2_DOMAIN_2; 2.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR SMART; SM00239; C2; 1.
SQ SEQUENCE 1399 AA; 158945 MW; 3F1550E3D52D237B CRC64;

Query Match 82.8%; Score 6856.5; DB 11; Length 1399;
Best Local Similarity 83.0%; Pred. No. 0;
Matches 1338; Conservative 17; Mismatches 22; Indels 235; Gaps 3;

QY 1 MSAPLGPGRPAPTPAASOPPOPEMDLSHLEERKIIIVAMDOKKEEKOQVYKI 60
Db 1 MSAPLGPGRPAPTPAASOPPOPEMDLSHLEERKIIIVAMDOKKEEKOQVYKI 59
QY 61 KEHKAQPTOWPFGSCITELVNVNVLPOOKOPNEKEPOTKLHOQFEYKQVYKMGESQ 120
Db 1114 KEHKAQPTOWPFGSCITELVNVNVLPOOKOPNEKEPOTKLHOQFEYKQVYKMGESQ 1173

```

```
Db 60 -----KLHQFEMYKEQVKMGESQ 80
OY 121 QOQEQKADATCGICHTKFKADGCGHNCVQCQTKPCARCGRVSLRSKKNVWVNCULCRQ 180
Db 81 QOQEQKADATCGICHTKFKADGCGHNCVQCQTKPCARCGRVSLRSKKNVWVNCULCRQ 140
OY 181 QELITKSGAMFYNSGNTLQOPDQKVPKGLNBEAPQEKAKLHEQPOFOGACGLSVA 240
Db 141 QELITKSGAMFYNSGNTLQOPDQKVPKGLNBEAPQEKAKLHEQPOFOGACGLSVA 200
OY 241 VEGRAHAGLRTOTIKNGSVKHQIASDMPDRKSPSVSDONRKYQSEEREDYSQV 300
Db 201 VEGRAHAGLRTOTIKNGSVKHQIASDMPDRKSPSVSDONRKYQSEEREDYSQV 260
OY 301 PSDGTMPSPSDYADRROSOPEPOFEYEGHLNYSRNSRGRHSKEYVDEDEVSREX 360
Db 261 PSDGTMPSPSDYADRROSOPEPOFEYEGHLNYSRNSRGRHSKEYVDEDEVSREX 320
OY 361 ERQREERYQARYRSDPNLARYPKPOPYEOMRIHAEVSRAREHRSIDVSLANAELED 420
Db 321 ERQREERYQARYRSDPNLARYPKPOPYEOMRIHAEVSRAREHRSIDVSLANAELED 380
OY 421 SRTSLRMBRPSQORSYSEERRAMENORSTSMERTRAQOGSSTPQRTSNHSPTIPRSP 480
Db 381 SRTSLRMBRPSQORSYSEERRAMENORSTSMERTRAQOGSSTPQRTSNHSPTIPRSP 440
OY 481 IPLDRPDMRADSLRKHQHLDPSSAVRKTREKMETMLRNDLSLSDOSESVRPPRRHK 540
Db 441 IPLDRPDMRADSLRKHQHLDPSSAVRKTREKMETMLRNDLSLSDOSESVRPPRRHK 500
OY 541 SKKGAMROVSLSSSEELASTPEYTSQDVELESESVSEKGSOKGRKTSQEOGLSDS 600
Db 501 SKKGAMROVSLSSSEELASTPEYTSQDVELESESVSEKGSOKGRKTSQEOGLSDS 560
OY 601 NTRSEOKRMVYGGSHLEEDLWEMSEPIKDSQVDTCSSTTLNBEHSHDKHPYTWQSK 660
Db 561 NTRSEOKRMVYGGSHLEEDLWEMSEPIKDSQVDTCSSTTLNBEHSHDKHPYTWQSK 620
OY 661 DGDRLIGRTLNRKLDGSPYRDSGAMGLKVVYGGKMTESGRLCFAITTKVYKGSILADYV 720
Db 621 DGDRLIGRTLNRKLDGSPYRDSGAMGLKVVYGGKMTESGRLCFAITTKVYKGSILADYV 680
OY 721 HLRPGDEVLEWNGRLIQGATFEEVYNILLESKPEQVELVSRPIDIPRIIPDSTHAOLE 780
Db 681 HLRPGDEVLEWNGRLIQGATFEEVYNILLESKPEQVELVSRPIDIPRIIPDSTHAOLE 740
OY 781 SSSSSFEESOKMRRPSISVSPMSPGMLRDVPOFLSGOLSTIKLWFDKVGHOLIVTILGAKD 840
Db 741 SSSSSFEESOKMRRPSISVSPMSPGMLRDVPOFLSGOLSTIKLWFDKVGHOLIVTILGAKD 800
OY 841 LPSREDGRPRNPVYKITYFLPDRSDKNKRTKTYAKTLEPKMNQOTFISPHKRREFRERML 900
Db 801 LPSREDGRPRNPVYKITYFLPDRSDKNKRTKTYAKTLEPKMNQOTFISPHKRREFRERML 860
OY 901 EITLWDQANVREESEFELEIILETALLDDEPHWYKLOTHDVSSLPLRPPSPYLPYLRQ 960
Db 861 EITLWDQANVREESEFELEIILETALLDDEPHWYKLOTHDVSSLPLRPPSPYLPYLRQ 920
OY 961 LHGESPTRLQSRKISDSSEVSYDCEDEGVGVYSDYRHNHRDLOSSLTSLSPVEQVSSNHC 1020
Db 921 LHGESPTRLQSRKISDSSEVSYDCEDEGVGVYSDYRHNHRDLOSSLTSLSPVEQVSSNHC 980
OY 1021 SPGSGPHRDVIGRTSRMSPSPAPPOANVBOGHRGTRATGHYNTISMDDRRHVMDDHYS 1080
Db 981 SPGSGPHRDVIGRTSRMSPSPAPPOANVBOGHRGTRATGHYNTISMDDRRHVMDDHYS 1040
OY 1081 DRD-----RDCEAADQRYHRSHTEQRPPLLERTTTSRSRSE 1118
Db 1041 ERDSHFLTLPRSRHROTSEHHNRDCEAADQRYHRSHTEQRPPLLERTTTSRSRSE 1100
OY 1119 PDTNLMRSPSLMTGSGAPSPALSLSHRRTGSGVOTSPSSPTPGTGRGRQOLPOLPRPGTL 1178
Db 1101 ADTNLMRSPSLMTGSGAPSPALSLSHRRTGSGVOTSPSSPTPGTGRGRQOLPOLPRPGTL 1160
```

```
OY 1179 ERSAMDIEERNROMKLNKYQOVAGSDPRLEQDYHSKTRSGMDPHRGADTVSTKSSDSQVS 1238
Db 1161 ER----- 1162
OY 1239 DVSAVRTSSASRFSSTSYMSVQSERPRGNRKISVTSKMNQNRQMGVSGKMLTSTISG 1298
Db 1163 ----- 1162
OY 1299 DMSLEKNDGSQSDTAVAGALGTSGKKRRSSIGAKMAIVGLSRKRSASOLSQTEGGGKK 1358
Db 1163 -----GGGKK 1167
OY 1359 LRSTVORSTETGLAVERNMNTROASRSTYDGSNYSSEGNLFFPGVRLADSQSPFL 1418
Db 1168 LRSTVORSTETGLAVERNMNTROASRSTYDGSNYSSEGNLFFPGVRLADSQSPFL 1227
OY 1419 DGLGPAQLVGRQTLATPAMGDIOVMMDKKQOLEVEITIRAGLVVKKPGSKTLPAIPYKVY 1478
Db 1228 DGLGPAQLVGRQTLATPAMGDIOVMMDKKQOLEVEITIRAGLVVKKPGSKTLPAIPYKVY 1287
OY 1479 LLDNGVCIAKKKTQVARKTLEPLYQOLLSEESPQGVLOIYWG DYGRMDHKSFMGVAQ 1538
Db 1288 LLDNGVCIAKKKTQVARKTLEPLYQOLLSEESPQGVLOIYWG DYGRMDHKSFMGVAQ 1347
OY 1539 ILDELELSNMVTGWFKLPPSSILVPTSAPLTRRASOSSLSTGSPYSRS 1590
Db 1348 ILDELELSNMVTGWFKLPPSSILVPTSAPLTRRASOSSLSTGSPYSRS 1399

RESULT 6
O9JIR7
AC O9JIR7 PRELIMINARY: PRT: 1330 AA.
DT 01-OCT-2000 (TREMBLrel, 15, Created)
DT 01-OCT-2000 (TREMBLrel, 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel, 16, Last annotation update)
DE RIM2-4C.
GN RIM2-4C.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347919; Pubmed=10748113;
RA Wang Y., Sugita S., Sudhof T.C.;
RT "The RIM/NIM Family of Neuronal C2 Domain Proteins. Interactions with
RT Rabs and a new class of Src homology 3 domain proteins."
RL J. Biol. Chem. 275:20033-20044(2000).
DR EMBL: AF199329; AAF81651.1; -.
DR InterPro: IPR000008; -.
DR InterPro: IPR000306; -.
DR InterPro: IPR000345; -.
DR InterPro: IPR001478; -.
DR Pfam: PF00168; C2; 2.
DR Pfam: PF00595; PDZ; 1.
DR PROSITE: PS50004; C2_DOMAIN_2; 2.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
DR SMART: SM00239; C2; 1.
DR SEQUENCE 1330 AA; 151044 MW; 485906E1F66FC73 CRC64;

Query Match 79.0%; Score 6543; DB 11; Length 1330;
Best Local Similarity 80.7%; Pred. No. 0;
Matches 1283; Conservative 18; Mismatches 29; Indels 260; Gaps 4;

OY 1 MSAPLPGRGAPPTPAASPPPOPEMDLSHLEERKIIILAVMDOKKEEKEQSVLKI 60
Db 1 MSAPLPGRGAPPTPAASPPPOPEMDLSHLEERKIIILAVMDOKKEEKEQSVLKI 59
OY 61 KEENKAOPTOMPFPFSGITELVNNVLIPOOKOPNEKEPOTKILHQOFEMYKEQVKMGESQ 120
Db 1101 KEENKAOPTOMPFPFSGITELVNNVLIPOOKOPNEKEPOTKILHQOFEMYKEQVKMGESQ 119
```

```

Db 60 ----- KLHQEFENYKREQVKKMGESQ 80
QY 121 QOOBOKGAPTCIGCHKTKEADGGCHNCSYCOTKPCARCGGRVSLRKNKMYCNLCRKQ 180
Db 81 QOOBOKGAPTCIGCHKTKEADGGCHNCSYCOTKPCARCGGRVSLRKNKMYCNLCRKQ 140
QY 181 QEILTKSGAFYNSGNTLQOPDQKVPRLRNEEAPQOEKKAKLHEOPQOFGAGDLSVPA 240
Db 141 QEILTKSGAFYNSGNTLQOPDQKALRLRSEAPQOEKKAKLHEOTOFOPGPGDSSVPA 200
QY 241 VEKGRAGLTPDQTKNGSGVKNQIASDMPDRKRSVSRDONRREYEOSEEREDYSQYV 300
Db 201 VERGRAGLTPDQTKNGSGVKNQIASDMPDRKRSVSRDONRREYEOSEEREDYSQYV 260
QY 301 PSDGMPSPSDYADRRSOREPOFEEPHLNYRDSNRGRHRSKTYVDEDEVESDEX 360
Db 261 PSDGMPSPSDYADRRSOREPOFEEPHLNYRDSNRGRHRSKTYVDEDEVESDEX 320
QY 361 ERORREERYQARYRSDPNLARYPVKPOPYEOMRIHAEVSRAHERHSDVSLAMELED 420
Db 321 ERORREERYQARYRSDPNLARYPVKPOPYEOMRIHAEVSRAHERHSDVSLAMELED 380
QY 421 SRISLLRMDRPSRQSVSEBRAMENORSYSMEKTRERAGOSSTYPORTNHSPTPRSP 480
Db 381 SRISLLRMDRPSRQSVSEBRAMENORSYSMEKTRERAGOSSTYPORTNHSPTPRSP 440
QY 481 IPLRDPOMRRADSLRKHHLDPSSAVRKTREKMETMLRNDLSLSSOSESVRPPPHK 540
Db 441 IPLRDPOMRRADSLRKHHLDPSSAVRKTREKMETMLRNDLSLSSOSESVRPPPHK 500
QY 541 SKKGKMRQVSLSSSEELASTPEYTSQDVELSESVSEKDGKGRKTSQGVLSDS 600
Db 501 SKKGKMRQVSLSSSEELASTPEYTSQDVELSESVSEKDGKGRKTSQGVLSDS 560
QY 601 NTRSEROKKRMVYGGHSLSEEDLEKSEPOIKDSGVDTCSTTLNEBHSKDKHPYTWQPSK 660
Db 560 NTRSEROKKRMVYGGHSLSEEDLEKSEPOIKDSGVDTCSTTLNEBHSKDKHPYTWQPSK 620
QY 661 DGDRLLIGRLILNKRLKDSVPRDSQGLKLVYGGKMTESGRCAFTTKVKKSLADTVG 720
Db 720 DGDRLLIGRLILNKRLKDSVPRDSQGLKLVYGGKMTESGRCAFTTKVKKSLADTVG 680
QY 721 HLRPDEVLEWNGRLLOGATEEVEVNIILSEKPEPOVELVSNRPIGDIRPIPOSTHAOLE 780
Db 680 HLRPDEVLEWNGRLLOGATEEVEVNIILSEKPEPOVELVSNRPIGDIRPIPOSTHAOLE 740
QY 781 SSSSFESQKMDRPSISVYSPMSPGMLRDVPOFLSGQLSIKLMFDKVGHOLIYTLIGAKD 840
Db 840 SSSSFESQKMDRPSISVYSPMSPGMLRDVPOFLSGQLSIKLMFDKVGHOLIYTLIGAKD 800
QY 841 LPSRDEGPRRNPVYIYFLPDRSDKNKRTKTVKTLEPKMNOTFIYSPVHRREFRRL 900
Db 900 LPSRDEGPRRNPVYIYFLPDRSDKNKRTKTVKTLEPKMNOTFIYSPVHRREFRRL 860
QY 901 EITLMDQARVRESEEFGEILIELEETALLDEPHWYKLTQTHVSSLPRLRPSYLLPRQ 960
Db 860 EITLMDQARVRESEEFGEILIELEETALLDEPHWYKLTQTHVSSLPRLRPSYLLPRQ 920
QY 961 LHGESPTRLORSKRISDSSEVDYDCGEGVGVSYRHNGRDLOQSTLSLPEQVMSSNHC 1020
Db 920 LHGESPTRLORSKRISDSSEVDYDCGEGVGVSYRHNGRDLOQSTLSLPEQVMSSNHC 980
QY 1021 SPGSGPHRVNDYIGRTSRMSPAPPOQANVEOGHGRGRATGHYNTISRMDDHRRVMDHYSS 1080
Db 980 SPGSGPHRVNDYIGRTSRMSPAPPOQANVEOGHGRGRATGHYNTISRMDDHRRVMDHYSS 940
QY 941 SPGSGPHRVNDYIGRTSRMSPAPPOQANVEOGHGRGRATGHYNTISRMDDHRRVMDHYSS 900
Db 900 SPGSGPHRVNDYIGRTSRMSPAPPOQANVEOGHGRGRATGHYNTISRMDDHRRVMDHYSS 860
QY 1081 DRDRDCEAADQPYHRSHSTEQRPPLLERTTTRSSSERPDNTLMRSMPSLMTGRSAPSP 1140
Db 1140 DRDRDCEAADQPYHRSHSTEQRPPLLERTTTRSSSERPDNTLMRSMPSLMTGRSAPSP 1100
QY 1141 ALSRSHPTGVSQVSPSSSTPGGRGRGROLPLPKKTLERSANDIEERNQMLNKTKQY 1200
Db 1200 ALSRSHPTGVSQVSPSSSTPGGRGRGROLPLPKKTLERSANDIEERNQMLNKTKQY 1160
QY 1201 AGSDRLBODYHSKYRSGMDPHRGADTVSGKSSSDSDVSDVSAVSRTSSASRFSSTYSMV 1260
Db 1094 ----- 1093
QY 1261 QSERPRGNKRKISVFTSKMONROMGVSGKNLTKSTISISGDCMSLEKNDGSDTAVALGAT 1320
Db 1094 ----- 1093
QY 1321 SGKRRSSIGAMVAIVGLSRKRSRASQLSQTEBGGKRLRSTYQVRSJETGLAEMRWMT 1380
Db 1094 ----- GGKRLRSTYQVRSJETGLAEMRWMT 1120
QY 1381 ROASRESTDGSNNYSSEGNLIFPGVRLASDSQFSPFLDGLGPAQLGROTLTAPAGDI 1440
Db 1120 ROASRESTDGSNNYSSEGNLIFPGVRLASDSQFSPFLDGLGPAQLGROTLTAPAGDI 1180
QY 1441 QVGMDDKQGLEVEIIRANGLVKPKSKTLPAPYVYLLDNGVCIAKKKTKVARKTLEP 1500
Db 1180 QVGMDDKQGLEVEIIRANGLVKPKSKTLPAPYVYLLDNGVCIAKKKTKVARKTLEP 1240
QY 1501 LYQQLSPFESPOGRVLOIITWGDYGRMDHKSFGVQAQIILDELESNNYIGFKLEPPS 1560
Db 1240 LYQQLSPFESPOGRVLOIITWGDYGRMDHKSFGVQAQIILDELESNNYIGFKLEPPS 1300
QY 1561 SLVDPTSAPLTRRASQSSLESSTGSPYSRS 1590
Db 1300 SLVDPTSAPLTRRASQSSLESSTGSPYSRS 1330

```

```

RESULT 7
Q9JHU6 ID Q9JHU6 PRELIMINARY; PRT: 1352 AA.
AC Q9JHU6.
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)
DE RIM2-5C (RIM2-2A) (RIM2-3B) (RIM2-4A).
GN RIM2-5C OR RIM2-2A OR RIM2-3B OR RIM2-4A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347919; PubMed=10748113;
KA Wang Y., Sugita S., Sudhof T.C.;
RT "The RIM/NIM Family of Neuronal C2 Domain Proteins. Interactions with
RT Rab3 and a new class of Src homology 3 domain proteins.";
RL J. Biol. Chem. 275:20034-20044(2000).
DR EMBL: AF199332; AAF81654.1; -
DR EMBL: AF199326; AAF81648.1; -
DR EMBL: AF199327; AAF81649.1; -
DR InterPro: IPR000008; -
DR InterPro: IPR000306; -
DR InterPro: IPR000345; -
DR InterPro: IPR001478; -
DR Pfam: PF00168; C2_2.
DR Pfam: PF00595; PDZ_1.
DR PROSITE: PS50004; C2-DOMAIN_2; 2.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
DR SMART: SM00239; C2_1.
SQ SEQUENCE 1352 AA; 153718 MW; 5E1E6F0E42DFEA9C CRC64;

```

Query Match 78.7%; Score 6522; DB 11; Length 1352;
 Best Local Similarity 79.6%; Pred. No. 0;
 Matches 1283; Conservative 18; Mismatches 29; Indels 282; Gaps 5;

```

QY 1 MSAPLGPGRAPAPTPAASPPQPEMPDLSHLTFEERKILLAVMDRKKEKEKQSVLKI 60
Db 1 MSAPLGPGRAPAPTPAASPPQPEMPDLSHLTFEERKILLAVMDRKKEKEKQSVLKI 59

```

```
OY 61 KEHKAQPTQWFPESGITELVNVLQPOQKQPNKEKQPOTKLHQFEMTKQEQYKKMGESQ 120
Db -----KTHQOFEMTKQEQYKKMGESQ 80
OY 121 QOQOBKQDAPTCGICHTKTFKADGCGHNCYCQTKFCARGCGHVSLRSNKMVMVNCILCRQ 180
Db QOQOBKQDAPTCGICHTKTFKADGCGHNCYCQTKFCARGCGHVSLRSNKMVMVNCILCRQ 140
OY 181 QELLTKGAMFYNGSNTLQOPDOKVPRGLNENAPQEKKAHLHOPQOGAGDLSPVA 240
Db QELLTKGAMFYNGSNTLQOPDOKVPRGLNENAPQEKKAHLHOPQOGAGDLSPVA 200
OY 241 VEGRAHGLTROPDTIKNGSGVKHQIASDMPDRKSPSVSDONRYQSEEREDYSQV 300
Db 201 VEGRAHGLTROPDTIKNGSGVKHQIASDMPDRKSPSVSDONRYQSEEREDYSQV 260
OY 301 PSDGTMRSPSDYADRRSQREPOFEPEPGLNYRDSNRGRHSKEYIYDDEDVSRDEY 360
Db 261 PSDGTMRSPSDYADRRSQREPOFEPEPGLNYRDSNRGRHSKEYIYDDEDVSRDEY 320
OY 361 ERGRREERYARISDNLAIRPYKPOPEEQMKIHAESRAHRRHSDVSLANAELLED 420
Db 321 ERGRREERYARISDNLAIRPYKPOPEEQMKIHAESRAHRRHSDVSLANAELLED 380
OY 421 SRTSLRMDPRRSOVSERRAAMENORSYMERTRRQOGOSYPORTSNHSPTPRRSP 480
Db 381 SRTSLRMDPRRSOVSERRAAMENORSYMERTRRQOGOSYPORTSNHSPTPRRSP 440
OY 481 IPLDRPMRRADSLRKQHLLDPSSAVRKTREKMETMLRNDLSLSDOSESVRPPRRPHK 540
Db 441 IPLDRPMRRADSLRKQHLLDPSSAVRKTREKMETMLRNDLSLSDOSESVRPPRRPHK 500
OY 541 SKKGKMRQVSLSSSEBELASTPEYTSQDVELESESEKGDQKGRKKSIBQCVLSDS 600
Db 501 SKKGKMRQVSLSSSEBELASTPEYTSQDVELESESEKGDQKGRKKSIBQCVLSDS 545
OY 601 NTRSEBOKRMYYGSHLEEDLEWSEPOIKDSGYDTCSTTLNEHSHSDKHPYTWQPSK 660
Db 546 -----YSMLHASMW-----SSEASMSLHPYTWQPSK 573
OY 661 DGDRLILGRILLNKLKLDGSPRDSGAMLGKLVGKMTESGRICAFITTKVKGSLADTVG 720
Db 574 DGDRLILGRILLNKLKLDGSPRDSGAMLGKLVGKMTESGRICAFITTKVKGSLADTVG 633
OY 721 HLRPGDEVLEMNGRLLOGATFEERYNTILSKPEPOVEIYVSRIGDIPRLPDSTHGLE 780
Db 634 HLRPGDEVLEMNGRLLOGATFEERYNTILSKPEPOVEIYVSRIGDIPRLPDSTHGLE 693
OY 781 SSSSFSSESQKMDRPSISVYSPMSPGMLRDVPOFLSGQLSIKLMPDKYGHQILVITLILAKD 840
Db 694 SSSSFSSESQKMDRPSISVYSPMSPGMLRDVPOFLSGQLSIKLMPDKYGHQILVITLILAKD 753
OY 841 LBSREGRGRNRPVYKTYFLPDRSDKDKRKRTKYKTKLEPKWNOTFIYSPVHRRFEREML 900
Db 754 LBSREGRGRNRPVYKTYFLPDRSDKDKRKRTKYKTKLEPKWNOTFIYSPVHRRFEREML 813
OY 901 ETLIMQARVRESESEFLGILTELETAALDDPWHYKLTOTHOVSSILPLRPSYTLRRQ 960
Db 814 ETLIMQARVRESESEFLGILTELETAALDDPWHYKLTOTHOVSSILPLRPSYTLRRQ 873
OY 961 LHGESFTRILQSRKSLSDSEVDYDCEDEGVGVSDYRHNHGRDLOQSTLSVPEQVMSSNHC 1020
Db 874 LHGESFTRILQSRKSLSDSEVDYDCEDEGVGVSDYRHNHGRDLOQSTLSVPEQVMSSNHC 933
OY 1021 SPSGSPHRYDVIGRTSMSPSPAPPPORANVEQGRGTRATGHYNTISMDRHRVMDHYSS 1080
Db 934 SPSGSPHRYDVIGRTSMSPSPAPPPORANVEQGRGTRATGHYNTISMDRHRVMDHYSS 993
OY 1081 DRD-----RDCEAADRQPYHRNSTORPLLETTTTRSSR 1118
Db 994 ERDSHFLTLPRSRKHROTSEHHNRDGDCCRAADKQPIHRKSTSEORPLLETTTTRSSR 1053
```

```
OY 1119 PDTNLKMSMPSLMTGRSAPPSPALSRSHPRGVSQVTSBSSSTPGTGRGRQOLPOLPKCTL 1178
Db 1054 ADYNLMKSMPSLMTGRSAPPSPALSRSHPRGVSQVTSBSSSTPGTGRGRQOLPOLPKCTL 1113
OY 1179 ERSAMDIEERNRQMKLKYQVAGSDPRLQODYHSKXRSBGMDPHRGADVTSTKSDSYVS 1238
Db 1114 ER----- 1115
OY 1239 DVSASRTSSASRFSSTSYMSVOSERPRGNRKISVFTSKMQRQMGVSGKMLTKSTSIG 1298
Db 1116 ----- 1115
OY 1299 DMSLEKNDGSDTAVGALGTSGKKRRSSITGAKMAIVGLSKRSASQLSQTEGGGKK 1358
Db 1116 -----GGGKK 1120
OY 1359 LRSTVOSTETGLAVEMRNMTROASRESTDGSNYSSEGNLTFPGVRLASDSQFDFL 1418
Db 1121 LRSTVOSTETGLAVEMRNMTROASRESTDGSNYSSEGNLTFPGVRLASDSQFDFL 1190
OY 1419 DGLGPAQLVGRQTLATPAMGDIOGMMDKKQLEVEIIRARGLVYKPSKTLPAIPYKVY 1478
Db 1181 DGLGPAQLVGRQTLATPAMGDIOGMMDKKQLEVEIIRARGLVYKPSKTLPAIPYKVY 1240
OY 1479 LLDNGVCIAKKTKVARKTLEPLXQULLSFEEBPQGRVLOIYWGDYGRMDHKSFGVQAQ 1538
Db 1241 LLDNGVCIAKKTKVARKTLEPLXQULLSFEEBPQGRVLOIYWGDYGRMDHKSFGVQAQ 1300
OY 1539 ILDELELENNVIGWFKLEPPSSLYDPTSAPLTRRSQSSLESSTGYSYSRS 1590
Db 1301 ILDELELENNVIGWFKLEPPSSLYDPTSAPLTRRSQSSLESSTGYSYSRS 1352
```

```
RESULT 8
O9JIR9 PRELIMINARY; PRT: 1368 AA.
AC O9JIR9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE RIM2-3A.
GN RIM2-3A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP
SEQUENCE FROM N.A.
RX MEDLINE=20347919; PubMed=10748113;
RA Wang Y., Sugita S., Sudhof T.C.;
RT "The RIM/MIM Family of Neuronal C2 Domain Proteins. Interactions with
Rab3 and a new class of Sic homology 3 domain proteins."
RL J. Biol. Chem. 275:20033-20044(2000).
DR EMBL: AF199325; AAF81647.1; -.
DR InterPro: IPR000008; -.
DR InterPro: IPR000306; -.
DR InterPro: IPR000345; -.
DR InterPro: IPR001478; -.
DR Pfam: PF00595; C2; 2.
DR ProSITE: PS50004; C2 DOMAIN_2; 2.
DR ProSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
DR SMART: SM00239; C2; 1.
SQ SEQUENCE 1368 AA: 155559 MW: CE08E16F0B383A11 CRC64;
```

```
Query Match 78.5%; Score 6504; DB 11; Length 1368;
Best Local Similarity 78.8%; Pred. No. 0;
Matches 1283; Conservative 18; Mismatches 29; Indels 296; Gaps 6;
OY 1 MSADLPGRGAPAPPAASQPPQPPMDLSHLEERKILILAVMDROKKEEKEQSVLKI 60
Db 1 MSADLPGRGAPAPPAASQPPQPPMDLSHLEERKILILAVMDROKKEEKEQSVLKI 59
```

```

QY 61 KEHKAQPTQWPFSGITELVNNVLQPOOKOPNEKEPQTKLHOQFENYKBOVKMCESEQ 120
Db 60 -----KLHOQFENYKBOVKMCESEQ 80
QY 121 QOQEQKGDAPTGICHTKTKPADCGCHNCYCOTKFCARCGGRVSLSNKVMVVCNLCRKQ 180
Db 81 QOQEQKGDAPTGICHTKTKPADCGCHNCYCOTKFCARCGGRVSLSNKVMVVCNLCRKQ 140
QY 181 QEILTKSAMWYNSGNTLQOPDQKVPBGLRNEEAQOEKAKLHEQDQPCGACGDLSPVA 240
Db 141 QEILTKSAMWYNSGNTLQOPDQKALRGLRSEEAQOEKAKLHEQDQFGPPGDDSSVA 200
QY 241 VEKGAHGLTRODTIKNSGVGHQIASDMPDRKRSPSVSDQNRREJOSEEREDYQOYV 300
Db 201 VERGAHGLTRODSTIKNSGKMHQIASDMPDRKRSPSVSDQNRREJOSEEREDYQOYV 260
QY 301 PSDGTMPSPSDYADRSQREPOFYEEPHLNYRDSNRGRHRSKEYIVDEDEVESRDEY 360
Db 261 PSDGTMPSPSDYADRSQREPOFYEEPHLNYRDSNRGRHRSKEYIVDEDEVESRDEY 320
QY 361 ERQREEEQOAYRSDPNLARYPVKQPYEEQMRHAEVSRARHERHSDVSLANALED 420
Db 321 ERQREEEQOAYRSDPNLARYPVKQPYEEQMRHAEVSRARHERHSDVSLANALED 380
QY 421 SRISLLRMDRPSRORSVSEBRAMENQSYSMERTREAOQSSYPORTSNHSPPTPRSP 480
Db 381 SRISLLRMDRPSRORSVSEBRAMENQSYSMERTREAOQSSYPORTSNHSPPTPRSP 440
QY 481 IPLDRPMRADSLRKQHLDPSSAVARKTKREKMETLRNDLSLSDQSESVPRPPPHK 540
Db 441 IPLDRPMRADSLRKQHLDPSSAVARKTKREKMETLRNDLSLSDQSESVPRPPPHK 500
QY 541 SKKGKMKOVSLSSSEELASTPEYTSQDVELESSEVSGEKDME----- 545
Db 501 SKKGKMKOVSLSSSEELASTPEYTSQDVELESSEVSGEKDME----- 545
QY 601 NTRSEQRKRMYYGHSLSLEJWSEPOIKDQSVDTCSSTLNEHSHSDKHPTWQPSK 660
Db 546 -----YSWLEHASWH-----SSEASPMGLHPTWQPSK 573
QY 661 DGDRLIGRILLNKLKDGSVPRDQSGMLGLKVYGGKMTESGRICAFTTKYKKGSLADTVG 720
Db 574 DGDRLIGRILLNKLKDGSVPRDQSGMLGLKVYGGKMTESGRICAFTTKYKKGSLADTVG 633
QY 721 HLRPDEVLENNGRLLQGTFFEEVYNIILLESKPEQVELVSRPIGIIPRIPDTHAOLE 780
Db 634 HLRPDEVLENNGRLLQGTFFEEVYNIILLESKPEQVELVSRPIGIIPRIPDTHAOLE 693
QY 781 SSSSFESQKMDRPSISVTPSPMSPGMLRDVPOPLSGOLS-----IKLMF 824
Db 694 SSSSFESQKMDRPSISVTPSPMSPGMLRDVPOPLSGOLS-----IKLMF 753
QY 825 DKVGHQILVTILGANDLPREDGPRPNRYVYIFLPRDSQDKNRKRTKYVKTLEPKMNQ 884
Db 754 DKVGHQILVTILGANDLPREDGPRPNRYVYIFLPRDSQDKNRKRTKYVKTLEPKMNQ 813
QY 885 FTYSVNHRRFEREMLEITLMDQARVRESESEFLGELLLELALDDEPHMYLQTHDV 944
Db 814 FTYSVNHRRFEREMLEITLMDQARVRESESEFLGELLLELALDDEPHMYLQTHDV 873
QY 945 SSLPRLRSPPYLPRRLOHGESPTRLQORSKRISDSEYSDVDCEDGVGVSDYRNGRDLO 1004
Db 874 SSLPRLRSPPYLPRRLOHGESPTRLQORSKRISDSEYSDVDCEDGVGVSDYRNGRDLO 933
QY 1005 SSTLSVPRQVWSSNHCSPGSPGPHRVYDVGRTSRWSBPAPORNVGEGHGRATRGHNT 1064
Db 934 SSTLSVPRQVWSSNHCSPGSPGPHRVYDVGRTSRWSBPAPORNVGEGHGRATRGHNT 993
QY 1065 ISRMDRHRVMDHYSRDR-----RCEADROPYHHSRSTEQ 1102
Db 994 ISRMDRHRVMDHYSRDR-----RCEADROPYHHSRSTEQ 1053

```

```

QY 1103 RPLERTTTRSSSRERPDNTNLMKSPSLMTGRSAPPSPALSRSHPRTGSVQTSPPSTPGT 1162
Db 1054 RPLERTTTRSSSRERPDNTNLMKSPSLMTGRSAPPSPALSRSHPRTGSVQTSPPSTPGT 1113
QY 1163 GRGRQOLPOLPKGTLEBRAMDIIEERNRQMKLNKYQVADSPLLEDYHNSKYSGMDPH 1222
Db 1114 GRGRQOLPOLPKGTLEBRAMDIIEERNRQMKLNKYQVADSPLLEDYHNSKYSGMDPH 1131
QY 1223 RGADIVSTKSSDSVDVSAVSRSTSSASRPSSTSYMSVQSERPRGNKISVFTSKMQNRQ 1282
Db 1132 ----- 1131
QY 1283 MGVSKNLTKSTISGDMCSLEKNDSQSDPTAVGALGTSGKRRSSIGAKWVAIVGLSRK 1342
Db 1132 ----- 1131
QY 1343 SRASASQSBTEGGKKLRSTVQSTETGLAVEKRNMTROASRSTGSMNSYSSEGLI 1402
Db 1132 -----GGKKLRSTVQSTETGLAVEKRNMTROASRSTGSMNSYSSEGLI 1180
QY 1403 PPGVRLASDQSFSDPLDGLGPAQLVGRQTLATPAMGDIQVGMMDKKQGLEVEIIRAGLV 1462
Db 1181 PPGVRLASDQSFSDPLDGLGPAQLVGRQTLATPAMGDIQVGMMDKKQGLEVEIIRAGLV 1240
QY 1463 YKPGSKTLPPAPYKVVYLLDNGVCIARKTKVARKTLEPLYQOLSFEEESPQGRVLOIYW 1532
Db 1241 YKPGSKTLPPAPYKVVYLLDNGVCIARKTKVARKTLEPLYQOLSFEEESPQGRVLOIYW 1300
QY 1532 GDYGRMDHKSFGVQAOIILDELELSMNVTGFKLFPSSLVDPISAPLITRASQSLSS 1582
Db 1301 GDYGRMDHKSFGVQAOIILDELELSMNVTGFKLFPSSLVDPISAPLITRASQSLSS 1360
QY 1583 TGPYSYRS 1590
Db 1361 TGPYSYRS 1368

```

RESULT 9

Q9JTR8 PRELIMINARY: PRT: 1292 AA.

AC Q9JTR8 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)

DE RIM2-4B.

GN RIM2-4B.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20347919; PubMed=10748113;

RA Wang Y., Sugita S., Sudhof T.C.;

RT "The RIM/NIM Family of Neuronal C2 Domain Proteins. Interactions with

R1 Rabi3 and a new class of Src homology 3 domain proteins.";

RL J. Biol. Chem. 275:20033-20044(2000).

DR EMBL: AF199328; AAF81650.1; -.

DR InterPro: IPR000008; -.

DR InterPro: IPR000306; -.

DR InterPro: IPR000345; -.

DR InterPro: IPR001478; -.

DR Pfam: PF00168; C2; 2.

DR Pfam: PF00595; PDZ; 1.

DR PROSITE: PSS0004; C2_DOMAIN_2; 2.

DR PROSITE: PSS00190; CYTOCHROME_C; UNKNOWN_1.

DR SMART: SM00239; C2; 1.

SQ SEQUENCE 1292 AA; 146933 MW; E15B2C2B25452094 CRC64;

Query Match 75.0%; Score 6212; DB 11; Length 1292;

Best Local Similarity 77.1%; Pred. No. 0;

Matches 1226; Conservative 21; Mismatches 45; Indels 298; Gaps 5;

```

OY 1 MSAPLPGRRPATPPASOPPOPEMDLSHLEERKILAVNDROKKEEKEOSVLKI 60
   |||||||
Db 1 MSAPLPGRRPATPPASOPPOPEMDLSHLEERKILAVNDROKKEEKEOSVLK- 59
OY 61 KEENKAOPTQWPFSGITELVNNVLOPOKOPNEKEBOTKLHQOFEMTKEOVKKKEESO 120
   |||||||
Db 60 -----KLHQOFEMTKEOVKKKEESO 80
OY 121 QOOEOGDAPTCICHTKTRKADCGHNCSTCYTKFCARCGGVSILRSKNVWVCNLCKKQ 180
   |||||||
Db 81 QOOEOGDAPTCICHTKTRKADCGHNCSTCYTKFCARCGGVSILRSKNVWVCNLCKKQ 140
OY 181 QELITSGAMFYSGSNTLQOPDOKVPRGLRNEEAPOEKKAKLHEOPFOGAGGLSLSPA 240
   |||||||
Db 141 QELITSGAMFYSGSNTLQOPDOKALGLRSEAPOEKKAKLHEOQTOFOQPPGDSVPA 200
OY 241 VERGRAHGLTRQDTIKNGSGVKHQIASDMPDRKRSFVSVDONRRRYEBOSEBEDYSQYV 300
   |||||||
Db 201 VERGRAHGLTRQDSIKNGSGMKHQIASDMPDRKRSFVSVDONRRRYEBOSEBEDYSQYV 260
OY 301 PSDGTPRBPSPDYADRRSQEPQFYEEPHLNTYDSDNRGRHSHKEYIVDEDEVSREY 360
   |||||||
Db 261 PSDGTPRBPSPDYADRRSQEPQFYEEPHLNTYDSDNRGRHSHKEYIVDEDEVSREY 320
OY 361 ERQOREEYQARYRSDPNLARYPVKPOPYEEOMRIHAESVSHRHRSDVSLAAELED 420
   |||||||
Db 321 ERQOREEYQARYRSDPNLARYPVKPOPYEEOMRIHAESVSHRHRSDVSLAAELED 380
OY 421 SRISLLRMDRPSQORSVSERRAAMENORSYSMERTREAOGOSYPQRTSNHSPPTPRSP 480
   |||||||
Db 381 SRISLLRMDRPSQORSVSERRAAMENORSYSMERTREAOGOSYPQRTSNHSPPTPRSP 440
OY 481 IPDPRDMRRAOISLRONHLDPSAVRKTREKMETLRLNDLSLSDOSESVSRPPRRPK 540
   |||||||
Db 441 IPDPRDLRRAOISLRONHLDPSAVRKTREKMETLRLNDLSLSDOSESVSRPPRRPK 500
OY 541 SKKGCMROVSLSSSEELASTPEYTSQCDVLELESSEVSSEKGSOKGKRKTSSEQVLSDS 600
   |||||||
Db 501 SKKGCMROVSLSSSEELASTPEYTSQCDVLELESSEVSSEKGSOKGKRKTSSEQVLSDS 545
OY 601 NTRSEROKRMATYGGHSLSEEDLEWSEPOIKDSGVDTCSSTTLNEHSHSKHAPVTWQPSK 660
   |||||||
Db 546 -----YSWLEHASMH-----SEASPMSLHPVTWQPSK 573
OY 661 DGDRLIGRILLNKRLKDSVPRRSGAMLGLKVYGGKMTESGRCAFTTKVKKGLADTVG 720
   |||||||
Db 574 DGDRLIGRILLNKRLKDSVPRRSGAMLGLKVYGGKMTESGRCAFTTKVKKGLADTVG 633
OY 721 HLRPGDEVLEWNGRLQOGATFEFEVYNILLESKPEPQVELVVSRIQDIPRIIPDSTHAOLE 780
   |||||||
Db 634 HLRPGDEVLEWNGRLQOGATFEFEVYNILLESKPEPQVELVVSRIQDIPRIIPDSTHAOLE 693
OY 781 SSSSSSEFSQKMDRPSISVSPMSPGMLRDVPOFLSGQLSIKLMWDKYGHOILVITLIGAKD 840
   |||||||
Db 694 SSSSSSEFSQKMDRPSISVSPMSPGMLRDVPOFLSGQLSIKLMWDKYGHOILVITLIGAKD 753
OY 841 LPSRDEGRBNPNPVKTYFLPDRSDKNKRRKTKYKTLLEPKMNQFFITSPIVRRRFRREML 900
   |||||||
Db 754 LPSRDEGRBNPNPVKTYFLPDRSDKNKRRKTKYKTLLEPKMNQFFITSPIVRRRFRREML 813
OY 901 ETLTMDQAVRESESEFLGILLETALDDDEPHWYKLOTHDVSSPLPPSPYTLPRQ 960
   |||||||
Db 814 ETLTMDQAVRESESEFLGILLETALDDDEPHWYKLOTHDVSSPLPPSPYTLPRQ 873
OY 961 LHGESPTRLQSRKSLSDSEVSDYDCEDEVGVVSDYRHNGRDLQSSSTLSVEQVWSSNHC 1020
   |||||||
Db 874 LHGESPTRLQSRKSLSDSEVSDYDCEDEVGVVSDYRHNGRDLQSSSTLSVEQVWSSNHC 933
OY 1021 SPGSGPHRDVIGRTSWSAPSAPPORANVQGHGRTATGHYNTISMDRHRVMDHYSS 1080
   |||||||
Db 934 SPGSGPHRDVIGRTSWSAPSAPPORANVQGHGRTATGHYNTISMDRHRVMDHYSS 993

```

```

OY 1081 DRRDCEAADROPYHRHRSTEQRPRLERTTTRSSRSSEPRDNTLKRSMPSLMTGSGAPSP 1140
   |||
Db 994 ERDSHFLTLPRSRHRQTSSEHHRD----- 1017
OY 1141 ALSRSHPRGTSVOTSPSSPTGTGRGROLPOLPKGLTERSAMDIEERNOMKLNKYQOV 1200
   |||||||
Db 1018 --GRSHPRGTSVOTSPSSPTGTGRGROLPOLPKGLTER----- 1055
OY 1201 AGSDPRLQDYHSKYRSGMPHKGADVTYIKSSDSVDVSAVSTSSASFSTSYMSV 1260
   |||||||
Db 1056 ----- 1055
OY 1261 QSEPRPGRNRIQVFTSKMQRQMGVSGKNLTKSTISGDMCSLEKNDGSQDFAVALGT 1320
   |||||||
Db 1056 -----GGKKLRSTVGRTSTETGLAVEERNMT 1082
OY 1321 SGKRRSSIGAKKVAIVGLSRKRSASOLSQTEGGKKLRSTVGRTSTETGLAVEERNMT 1380
   |||||||
Db 1056 -----GGKKLRSTVGRTSTETGLAVEERNMT 1082
OY 1381 RQASRESTGSMNSYSSEGNLIFPGVRLASDSQPSDFLDGLGPAQLVGRQTLATPAMGI 1440
   |||||||
Db 1083 RQASRESTGSMNSYSSEGNLIFPGVRLASDSQPSDFLDGLGPAQLVGRQTLATPAMGI 1142
OY 1441 QVGMMDKQGLEVEIIRARGLVYKPGSKTLPAPYKYVLLDNGVCIAKKTKVARKTLEP 1500
   |||||||
Db 1143 QVGMMDKQGLEVEIIRARGLVYKPGSKTLPAPYKYVLLDNGVCIAKKTKVARKTLEP 1202
OY 1501 LYQQLISFEESPOGRVLOITVWDYGRMDKSPFGVAQIILDELELSNMVIGMKFLPPS 1560
   |||||||
Db 1203 LYQQLISFEESPOGRVLOITVWDYGRMDKSPFGVAQIILDELELSNMVIGMKFLPPS 1262
OY 1561 SLVDPTAPLTRRASQSSLESSTGPSYSRS 1590
   |||||||
Db 1263 SLVDPTAPLTRRASQSSLESSTGPSYSRS 1292

```

```

RESULT 10
090026 PRELIMINARY; PRT; 1188 AA.
AC 090026;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE KIA00751 PROTEIN.
GN KIA00751.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=9087487; PubMed=9872452;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT For large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
DR EMBL; AB018294; BAA34471.1; -.
DR HSSP; P31016; 1BE9.
DR InterPro; IPR000008; -.
DR InterPro; IPR001478; -.
DR Pfam; PF00168; C2; 2.
DR Pfam; PF00595; PD2; 1.
DR PROSITE; PS50004; C2 DOMAIN_2; 2.
DR SMART; SM00239; C2; 1.
SQ SEQUENCE 1188 AA; 135249 MW; DD65DB32288055AF CAC64;

```

Query Match 68.1%; Score 5638; DB 4; Length 1188;
 Best Local Similarity 83.3%; Pred. No. 0;

```
Matches 1099; Conservative 24; Mismatches 23; Indels 174; Gaps 3;
OY 273 RKRSPEVSDONRRYEDSEEREDYSQYVPSDGTMPSPSDYADRSDRQPOFEFEGHLN 332
Db 41 RKNSPVSIRDONRRKIDQREEREYSQYATSDTAMPRSSDVIADRNSQHEQYETEDDHLN 100
OY 333 YRDSNRKGRHRSKEYIVDEDEDVESHDEYERQREBEYOARYRSDPNLARYPYKPPQYEEQ 392
Db 101 YRDSNRKGRHRSKEYIVDEDEDVESHDEYERQREBEYOARYRSDPNLARYPYKPPQYEEQ 160
OY 393 MRIHAEVSRAHRRHSDVSLANAELEDRSISLMDRPSRORSYSERRAAMENQSYSM 452
Db 161 MRIHAEVSRAHRRHSDVSLANAELEDRSISLMDRPSRORSYSERRAAMENQSYSM 220
OY 453 ERTEAGQSSYPTRTSNHSPPTPRSPPLDPRPMRADSRKONHLDPSSAVRKTRE 512
Db 221 ERTEAGQSSYPTRTSNHSPPTPRSPPLDPRPMRADSRKONHLDPSSAVRKTRE 260
OY 513 KMETMLRNDLSLSDQSESVRPPPRPHKSKKGKMRQVSLSSSEELASTPEYTCDDVE 572
Db 281 KMETMLRNDLSLSDQSESVRPPPRPHKSKKGKMRQVSLSSSEELASTPEYTCDDVE 340
OY 573 LESHVSVEKDSQCKKRTSQCVLSDSNTRSEKOKKRYTGGHSLSEEDLEWSEPOIKDS 632
Db 341 IESHSVEKDSQCKKRTSQCVLSDSNTRSEKOKKRYTGGHSLSEEDLEWSEPOIKDS 400
OY 633 GVDTCSSITLNEESHSHKHVYTMOPSKDGRILIGRILINKLKDGSVPRDSGAMGLKY 692
Db 401 GVDTCSSITLNEESHSHKHVYTMOPSKDGRILIGRILINKLKDGSVPRDSGAMGLKY 460
OY 693 VGGKMTESGRICAFITTKYKKSGLADTVGHLRPGDEVLEWNGRLQGAATEEVYNIILESK 752
Db 461 VGGKMTESGRICAFITTKYKKSGLADTVGHLRPGDEVLEWNGRLQGAATEEVYNIILESK 520
OY 753 PERQVELVSPRTGDIPIPIPDSTNAQLSSSSSFSSQKMDRPSISVTSFMSPGMLRDVPQ 812
Db 521 PERQVELVSPRTGDIPIPIPDSTNAQLSSSSSFSSQKMDRPSISVTSFMSPGMLRDVPQ 580
OY 813 FLSQLSLKLMFDDVGHOLIYTIIGAKLPREDGRPRNPYKIYFLPDRSDKNKRTKT 872
Db 581 FLSQLSLKLMFDDVGHOLIYTIIGAKLPREDGRPRNPYKIYFLPDRSDKNKRTKT 640
OY 873 VKKTLERKMNQTFIYSPVHREFFERMLEITLMDQARVREESEFGEILIELETAALLD 932
Db 641 VKKTLERKMNQTFIYSPVHREFFERMLEITLMDQARVREESEFGEILIELETAALLD 700
OY 933 EPHMYKIQTHDVSSLPRLRSPRYLPRROLDHGESPTRRLQSRKISDSEVSDVDCEDGVGY 992
Db 701 EPHMYKIQTHDVSSLPRLRSPRYLPRROLDHGESPTRRLQSRKISDSEVSDVDCEDGVGY 760
OY 993 VSDVRHNGRDLQSSSTLSVPEQVMSNHCSGSPGPHRVDIIGTRSMSPSPAPRPO -BNVEQ 1051
Db 761 VSDVRHNGRDLQSSSTLSVPEQVMSNHCSGSPGPHRVDIIGTRSMSPSPAPRPO 820
OY 1052 GHRCSTRA -TGHYNTISRDHRRVMDHYSDDRDCEADADROPYHNSRSTEOPLLETT 1110
Db 821 GLRGTFRMTGNYNTISRDHRRVMDHYSDDRDCEADADROPYHNSRSTEOPLLETT 880
OY 1111 TRRSSESRPTNLKRSMPSLMTGSAFPSALSRSHPRTGVSQVTSFSSPTGTRGRGRLP 1170
Db 881 TRRSSESRPTNLKRSMPSLMTGSAFPSALSRSHPRTGVSQVTSFSSPTGTRGRGRLP 940
OY 1171 QLPKGTGLERSAMDIEERNQMKLNKYQVAGSDPRLBDYHSKYSGMDPHNGCAVTST 1230
Db 941 QLPKGTGLERSAMDIEERNQMKLNKYQVAGSDPRLBDYHSKYSGMDPHNGCAVTST 952
OY 1231 KSSSDSVSDVSAVSRSTSASRFSSTSYMSVOSERPRGNRKISVFTSKMONRQMGVSGKNL 1290
Db 953 ----- 952
OY 1291 TKSTISIGDMCSLEKNDGOSDSTAVALGTSGKRRRSSIGAKMVAIVGLSRKRSASOLS 1350
Db 953 ----- 952
```

```
OY 1351 QTEGGGKILRSTVQSTETGLAVEMRNMMTROASRESTDGSNMSYSEGNLIFPGVRLAS 1410
Db 953 ----GGRKLRSIVQSTETGLAVEMRNMMTROASRESTDGSNMSYSEGNLIFPGVRLAS 1008
OY 1411 DSQFSDFLDGLGPAOLVGRQTLATPAMGDIQVGMMDKQOLEVEIIRARGLVYKPGSKTL 1470
Db 1009 DSQFSDFLDGLGPAOLVGRQTLATPAMGDIQVGMMDKQOLEVEIIRARGLVYKPGSKTL 1068
OY 1471 PAPYKVVLLDNGVCIAKKKTIVARKTLEPLVQQLSFESSPGGRVLIQIIVMGDYGMDH 1530
Db 1069 PAPYKVVLLDNGVCIAKKKTIVARKTLEPLVQQLSFESSPGGRVLIQIIVMGDYGMDH 1128
OY 1531 KSPFMGAQIILDELELSNMVIGMFLEPPSSLVDPPTSAPLTRPASQSLSTGSPYSRS 1590
Db 1129 KSPFMGAQIILDELELSNMVIGMFLEPPSSLVDPPTSAPLTRPASQSLSTGSPYSRS 1188

RESULT 11
O9JIR4
ID O9JIR4 PRELIMINARY: PRT: 1615 AA.
AC O9JIR4:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE RIM1B.
GN RIM1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347919; PubMed=10748113;
RA Wang Y., Sugita S., Sudhof T.C.;
RT "The RIM/NIM Family of Neuronal C2 Domain Proteins. Interactions with
RT Ratz3 and a new class of Src homology 3 domain proteins.";
RL J. Biol. Chem. 275:2003-2004(2000).
DR EMBL: AF199333; AAF81655.1; -.
DR InterPro: IPR000008; -.
DR InterPro: IPR000194; -.
DR InterPro: IPR000306; -.
DR InterPro: IPR000345; -.
DR InterPro: IPR001360; -.
DR InterPro: IPR001478; -.
DR Pfam: PF00168; C2; 2.
DR Pfam: PF00595; PDZ; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR PROSITE: PS50004; C2_DOMAIN_2; 2.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE: PS00572; GLYCOSYL_HYDROL_F1_1; UNKNOWN_1.
DR SMART: SM00239; C2; 1.
SQ SEQUENCE 1615 AA; 179653 MW; 80E76F74BF35FB7E CRC64;
```

Query Match 54.5%; Score 4513.5; DB 11; Length 1615;
Best Local Similarity 55.7%; Pred No. 2.4e-289;
Matches 961; Conservative 208; Mismatches 312; Indels 243; Gaps 37;

```
OY 1 MSAPLGRGPAFTPAASQPPQPEMPDLSHLTEERKILAVMDROKKEBEKQSVLKI 60
Db 1 MSSAVGPRGPRPT----VPPMQELPDLSHLTEERNITIMAVMDRQKEBEKEEAMLK 56
OY 61 KEENKAQ-----TQWFPSGITELYNNYLPQOKPNEKEPQT -KLHQFEMYKE 110
Db 57 VVRDMAKPAACKTPRNAESQPHQPPPLNIFRCVCPPRKPSSEEGGPERDMRLHQOFESYKE 116
OY 111 QVKKMGESQOOQ -EOKGAPTCGICHTKTFADGCGHNSCYQTKFCARCAGRGVLSLNK 169
Db 117 QVRKIGEFARRYQGEHKDAPTCGICHTKTFADGCGHLSYCTKTCARCAGRGVLSLN 176
OY 170 ---VMWVCNLCKRQOEILTKSGAMFYNSGSNTLQOPDQKVPGLRNE-----EAPQE 218
Db 170 ---VMWVCNLCKRQOEILTKSGAMFYNSGSNTLQOPDQKVPGLRNE-----EAPQE 218
```

```

Db 177 EDKVVWVWVNCNCRKQOELLTKSGAMFGSCP---QOPSOD---GTLSDTATGAGSEVPRE 230
QY 219 KKAKLHPOPOGAPGLSPVAVEKGRAHGLTROTITKNGS-----GKXH 263
Db 231 KKAKLOERSHSQ---TPLSTAAY-----SSQDTATPGAPLHRNKGAEPSQALPQEQ 279
QY 264 QIA-----SDMPSRKRSRPSRDONRRYEQSEBEDYSQYV--PSDGTMPRSPDYADR 317
Db 280 KQASRSRSEPRRKRKAPGLS-EONGKGQKSEKRVKSVQDGEI-----ADERERK 333
QY 318 SOEPQFEEPCGLNTRDSNRGRHSKEYIVDEVEDVESHEVERQREERYOARYSDP 377
Db 334 ERRETRLEKGRSQDYSD--RPEKRDNGRAVD-----QKQKEEYQTRYSDP 381
QY 378 NLARYPVKPOPYEORHIAEVARHERHSHVSLANALEDSRLSLMDRPSROSV 437
Db 382 NLARYPVKAPPEOQORHMAHVARHHRHSHVAPLPHNE-----AAAAAP 427
QY 438 SERRAAMONRSTSMERTREAOQSSYPQRTSNHSPPTPRSPPLIDRPDMRADSLRQ 497
Db 428 ABAATGKRAPATAVARVSPSPRARAAAPPTPHGPPRRPACRAPEPEPRVPEPLRQ 487
QY 498 HLLDPSAY--RKTREKMETMLRNDLSLSDQSESVRPPRPKSKGKMRQVSLSS 555
Db 488 GLDLDGSAVILKAKAREKESMLRNDLSLSDQSESVRPPRPKRGKRGKRVSS 547
QY 556 EEBLASTPEYTCDDVELESESVSEKGSQKGRKRTSEQVLSDSNTRSEKQKRYVGG 615
Db 548 EEEGVSTPEYTCDDVELESESVSEKGDLD-----YY-- 579
QY 616 HSLBEDLPMSEPOINDSGVDCSTTLNEHSHSKHPVTWQPSKDGRLIGRIILLNKL 675
Db 580 -----WLDP-----ATWHSRETSPISSHPTWQPSKEGDRLLIGRIILLNKT 620
QY 676 KDGVPKDSGAMGLKVVYGGKMTESGRCAPFTTKVKGLADVTGHLRPGDVLWNGRL 735
Db 621 ---TMPKESGALLGLKVVYGGKMTDLGRGAFITTKVKKSLADVGHGLRAGDVLWNGRP 677
QY 736 LOGATPEEVYVITLESREPEYELVSPRIDIPRIDSTHAULESSSSSEFSQKMDRPS 795
Db 678 LPGATNEEVYVITLESSEPEVEIIVSRPIDIPRIPESHHPLESSESSSEFSQKMDRPS 737
QY 796 ISVTSFMSPGMLRDVPOFLSGOLSIKLMFDKVGHOLYITIGADLPBREGCRPNPYV 855
Db 738 ISVTSFMSPGMLKADPVLPQOLSVKLMYDKVGHOLYVNLQADLPBRYVGRPNPYV 797
QY 856 IYFLPDSDNKRTKTVKTLBPKMNOTLYSPVHRERERMLETITLMQAVRPEES 915
Db 798 MYFLPDSDSKRTKTVKTLBPKMNOTLYSHVHRDRFERMLETITLMQAVRPEES 857
QY 916 EPLGELLIELETAALDDPEMYKLOTHDVSLPLRPSPLLRQQLGSESTRLORSKR 975
Db 858 EPLGELLIELETAALDDPEMYKLOTHDVSLPLRPSPLLRQQLGSESTRLORSKR 917
QY 976 ISDSSEVDICEDGVGVV--DYRHNGRDLOSSLTLYPEQVMSNHCSPSGPHRYVIG 1033
Db 918 ISDSISDSYEDVDIGVGVVYRASAAREKATTLTVPEQORTTHHSRVSYHRKGDQ 977
QY 1034 RTFSMSPSAPPPORANVEOGH--RGTRATGHYNTISRM---DRHRVMDHYSSD----- 1081
Db 978 RFRSRPLPNV-PLORSIDELHPTFRSRSPTRHHDASRSPADHRSRHVESQSSPDSLELM 1036
QY 1082 -----RDRDCEA----- 1088
Db 1037 LPRAKRGSAEESLHMTSELQPSLDRARASATNCLRPDTLSHSPERHRHSKSECSIOQO 1096
QY 1089 -----AD-----KQYHRSRSTQRPRLERTTTRSSSERPDTNLMRSMPLMT 1132
Db 1097 SRKGTASDADRVLPCLSRGYATPATKQDPVYRGKYPTRSRSSSEHSYRTLCSMHMLAP 1156
QY 1133 GRASAPSP-ALSRSHPRGTSVOTSSPTGCGRGOLPOLPK-GTLERSAMLDIEENR 1190
Db 1157 GGSAPSPULLLIRHTRQSPVQSPPADTISFGSRGRGOLPQVYVRSISIOASIVVEEFT 1216

```

```

QY 1191 QM--KLNTKYOV--AGSDPRLIEDYHNSKYRSGWDPHRRGADVTYTKSSDSDYDVAVSRT 1246
Db 1217 QMKYVHRFKQJOTSSGSSQOELDHEQSKYNIHKKDQYRSCDNAMAKSSSDSDYDVAISRA 1276
QY 1247 SSARFSSSTSYMSVQSEPRRNRKRTISVFTSKMQRNOKMGVSKNLTSTISGDMCSEKN 1306
Db 1277 SSTRLSTSTMSQOSERPG--RISFTPMQGRMGTSGRAIITKSTYSGEIYTLERN 1334
QY 1307 DGSQSDTAVALGATGCKRRSSISGAKMVAIYGLSRKSRASASLOTGGGCKLSTVQRS 1366
Db 1335 DGSQSDTAVALGATGAGCKRRSSISAKVAIY--SRRSRSTQSLOTSGHKKLSTIORS 1392
QY 1367 TENGIAVENMNMTRQASRSTDSNMSYSSEGNLIPPGVRLASQPSFLDGLGPAOL 1426
Db 1393 TENGMAEMKR--MKQPSRSTDSINSYSSEGNLIPPGVAVGDPDSQPSFLDGLGPAOL 1451
QY 1427 VGRQTLATPAMGDIQVGMMDKQOLEVEIIRARGLVYKPSKTLPPARYVYLLDNVCI 1486
Db 1452 VGRQTLATPAMGDIQGMEDKQOLEVEIIRARSLTQKPSKSTPARYVYVYLENGACT 1511
QY 1487 AKKTKVARKTLEPLVQQLSFEESPOGRVLOITVWGDYGRMDHKSFMGVAQILLDELEL 1546
Db 1512 AKKTKVARKTLDPLVQQLSVPFDESPOGKVLQYIWMGDYGRMDHCKFMGVAQILLDELD 1571
QY 1547 SNNVIGMFKLPPSSLVDPSTAPLTRRASQSSLESSTGSPSRS 1590
Db 1572 SNNVIGMFKLPPSSLVDPSTAPLTRRASQSSLESSTGSPSRS 1615

RESULT 12
035168 PRELIMINARY; PRT; 1553 AA.
ID 035168
AC 035168;
DT 01-JAN-1998 (TREMBLrel. 05. Created)
DE 01-JAN-1998 (TREMBLrel. 05. Last sequence update)
DR 01-MAR-2001 (TREMBLrel. 16. Last annotation update)
DE RAB3 EFFECTOR.
GN RAB3
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=97394473; PubMed=9252191;
RA Wang Y., Okamoto M., Schmitz F., Hofmann K., Sudhof T.C.;
RT "Rim is a putative Rab3 effector in regulating synaptic-vesicle
fusion."
RL Nature 388:593-598(1997).
DR EMBL: A9007836; AAB66703.1; -.
DR HSSP: P21707; IRSY.
DR InterPro: IPR000008; -.
DR InterPro: IPR000194; -.
DR InterPro: IPR000306; -.
DR InterPro: IPR001360; -.
DR InterPro: IPR001478; -.
DR InterPro: IPR003315; -.
DR Pfam: PF00168; C2; 2.
DR Pfam: PF00595; PD2; 1.
DR Pfam: PF02318; RPH3A_effector; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; UNKNOWN.1.
DR PROSITE: PS00004; C2_DOMAIN_2; 2.
DR SMART: SM00239; C2; 1.
SQ SEQUENCE 1553 AA; 172964 MW; D664397E65D03BE CRC64;

```

Query Match 54.2%; Score 4489.5; DB 11; Length 1553;
 Best Local Similarity 57.2%; Pred. No. 8.8e-288;
 Matches 952; Conservative 209; Mismatches 319; Indels 183; Gaps 37;

```

OY 1 MSAPLPGRRPAAPPAASQPPPOPEMDLSHLEERKILLAVMDROKKEEKEQSVLKI 60
Db 1 MSSAVGPPGRPRPT-----VPPMOELPDLSHLLEERNITIMAVMDROKKEEKEEAMLK 56
OY 61 KEENAO-----TQFPFSGITELVNNVLOPOOKOPNEKEPOT--KLHOOFEMIK 110
Db 57 VVRDMAKPAACKTPRNASQPHQPLNIFRCVCPRRKPSSEEGGPEMDRLHOOFESYKE 116
OY 111 QVKKMGESOOOQ--ROKGDAPTCGICHTKTRPAGCGHNCSCYOTKFCARCGHVSLSNK 169
Db 117 QVKTIGEARRKQGHGKHDDAPTCGICHTKTRPADCGHLSYCTKTCARCGHVSLSNN 176
OY 170 ---VMVNCNCRKQOELLTKSGAMFYNGSNTLOPODKVPRGLRNE-----EAPOE 218
Db 177 EDKVVWVNCNCRKQOELLTKSGAMFGSGP---QOPSOD---GTLSDTATGAGSEVPRE 230
OY 219 KKAKEHQPQPOGAGDLSVPAVEKGRANGLTRQDTIKGS-----GVKH 263
Db 221 KKAQLQERSRQ---TPISTAAY-----SSQDTATPGAPLHRNKGAPESQALGPQ 279
OY 264 QIA-----SDMPDRKRSPSVSDONRRYEOSFEREDYSQV--PSDGTMRSPSDVADR 317
Db 280 KOASRSRSEPPREKKAPOLS--EQNGKGQKSERKRPKSVYOPGEGT-----ADERERK 333
OY 318 SQREPOFYEEPHLNYRDSNRGRHRSKEYIVDEDEVESKDEXERQREBEYOARYSDP 377
Db 334 ERRETRRLKEGRSQYSD---RPEKRDNGRVAED-----QKQKEEYQTRYSDP 381
OY 378 NLARKPVAPPOPEEOMRIHAESRAHERHSDVSLANAELEDNRSLRLMDRPSRORSV 437
Db 382 NLARPVAPPEEQOMRHARVSRARHERHSDVALPHE-----AAAAAP 427
OY 438 SERRAAMENORSYMERTRPAGOSYSYPTQTSNHSPTPRSPILDRPMRADSRKQ 497
Db 428 AEATAGKAPATARKSPSPSPARAAAAPRTHEGPPPPRPAAGPAPPPRPPRPELRKQ 487
OY 498 HHLDPSAV--RKTREKMETMLRNDLSLSDOSEVVRPPRPHKSKGKMRQVSLSS 555
Db 488 GRLDGSAVLLRKAREKESMLRNDLSLSDQESVVRPPRPHRPRGCKRRQMSVSS 547
OY 556 EEELASTPEYITSCDDVLESESVSEKDGSKRKTSEQCVLSDNTRSRQKRMYYG 615
Db 548 EEEGVSTPEYITSCDEVELESESVSEKGDLD-----YY-- 579
OY 616 HSLERDLMSPEPQKDSGVDTCSSTTLENEHSHDKPVTWQPSKODRLIGRILLKRL 675
Db 580 -----WLDP-----ATWHSRETSPISHPTWQPSKEDRLIGRYILLKRT 620
OY 676 KDGSVPRDSGAMLGKVVYGGKMTESGRLCAFTTKVKKGSLADTVGHLRPGDEVLEWNGRL 735
Db 621 ---TPMKESGALLGLKVVYGGKMTDLGRGALFTTKVKKGSLADVVGHLRAGDEVLEWNGRP 677
OY 736 LQGAFFEEVYNTILSKSPROYELVYSRIGDIPRIPDSTHAOLESSSSSFESQKMDRPS 795
Db 678 LPGAINEEVYNTILSKSPROYEIVSRIGDIPRIPESSHAPLESSESSSFESQKMRPS 737
OY 796 ISVTPSPMGMLRDVPOFLSGOLSTIKLWDFYGHOLIVITLGLKODLPSRDEGPRNRYVK 855
Db 738 ISVTSPTSPGALKADAPQVLPGLSGLKWDKYGHOILIVANVLOATDLPKPRDGRPRNRYVK 797
OY 856 IYPLPDRSDKKRRTKTVYKTLPEKWNQTFIYSPVHRRFERREMLETTLMDQARVEES 915
Db 798 MYPLPDRSDKKRRTKTVYKTLPEKWNQTFYVSHVHRDREREMLETTVMDQARVOEES 857
OY 916 EFLGEIILELTALLDDPRHWKLOTHTVSSLPPLRPSYLPKROLHGESPTRLQSKR 975
Db 858 EFLGEIILELTALLDDPRHWKLOTHTHSSSLPLPQSPFMPRRRHGESSSKKLOHRSQR 917
OY 976 ISDSEVSDYDCEDGVYVS--DYRHNGRDLQSTLSVPCVMSNNKSPGSPHRAVYIG 1033
Db 918 ISDSDISDYEVDGIGVYPRVGYRASARSKATITLTVPEQQTTHHRSKVSFHRGDDQG 977
OY 1034 RTRSNWSPSAPPQORNVQGH--RGTRATGHTYNTISRM---DRHRVMDHYSDDRDRDCEA 1088

```

```

Db 978 RRRSLPWN--PIQRSDEIHTPTRRSRSPTRHHNDASSPADHRNRHVSQXSEPPSELML 1036
OY 1089 ADROPTNRS-----RTEORPLLERTTTRSRSEBRDNL---MRSPSLMT----- 1132
Db 1037 LPRARRGRSAESLHMTSELOPISLDRARSASTNCLRPTSLASPERERHRSKSERCSIOKQ 1096
OY 1133 GRSAPPSPALSRSHPRGTSVQTSPPSGTGRGRGROLPOLPPK--GTLERAMDIEERNQ 1191
Db 1097 SKKGTASDA--DKTHROGSTQSPADTSGFRGRGROLPOVPVYVSGSTLEQASLYVEETRQ 1155
OY 1192 M--KLNKYQV--AGSDPRLEODYHNSKYSRGWDPHRRGADTVSTKSSDSVDVSAVSRTS 1247
Db 1156 MKVYHRRFQOTGGSSQGLDHBQYSKYNIHKDOYRSCDNASAKSSPDSVDVSAISRAS 1215
OY 1248 SASRFSSTSYMSVOSRPRGNKRKISVFTSMONROKCVSGSKNLTKSTISGDMCSLEKD 1307
Db 1216 STSRLSSTSFMSQSERPRG--RISSETPRMOGRRMGTSGRAIKTSYSGEITLERND 1273
OY 1308 GSQSDTAVALGTSGKKRRSSIGAKMVAIYGLSRKRSASASOLQTEGGGKKIRSTVOGST 1367
Db 1274 GSQSDTAVALGTAVAGCKKRRSSLSAKYVAIV--SRSRSTLSQTESGHHKTKSTIORST 1331
OY 1368 ETGLAVEMRNMTROASRSTDSGSMNSYSSEGNLIPGVRLASDSQSFDELGLGPAQV 1427
Db 1332 ETGMAEMRK--MYRQPSRSTDSGINSYSSEGNLIPGVRLVGVDSQSFDELGLGPAQV 1390
OY 1428 GROTATLAPMAGDIQVGMAMKQGLEVEIIRARGLVYKPGSKTLPAPVYKYLLDNGYCIA 1487
Db 1391 GROTATLAPMAGDIQIGEMEKQGLEVEIRARSLTQKPGSKTPAPVYKYLLLENGACIA 1450
OY 1488 KKTIVARKTEPLLYQOLSFEEPSQGRVLOLYTWGIDYGMHDHKSFGVNOILLDELELS 1547
Db 1451 KKTITARTKTLDPLOOSLVFDESPOGKYVQVLYWGDYGMHDHCKFPNGVNOILLERDLS 1510
OY 1548 NNVIGWFKLEPPSSLVDPSTAPLTRRASQSLSSSTGSPSYSRS 1590
Db 1511 SWVIGWYKLEPPSSLVDPTRAPLTRRASQSLSSSGPPCIRS 1553
RESULT 13
O15048 ID 015048 PRELIMINARY; PRT; 1053 AA.
AC 015048;
AD 015048;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE KIA0340 (FRAGMENT).
GN KIA0340.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N. A.
RC TISSUE=BRAIN;
RX MEDLINE=9734984; PubMed=9205841;
RA Tanaka T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
DR EMBL; AB002338; BAA20798.1; -.
DR HSSP; P21707; IRSY.
DR InterPro; IPR000008; -.
DR InterPro; IPR000306; -.
DR InterPro; IPR001478; -.
DR InterPro; IPR001565; -.
DR InterPro; IPR003151; -.
DR Pfam; PF00168; C2, 1.
DR Pfam; PF00595; PDZ, 1.
DR Pfam; PF02318; RPH3A_effector; 1.

```

DR PRINTS: PR00399; SYNAPTOTAGM.
DR PROSITE: PS50004; C2_DOMAIN_2; 1.
DR SMART: SM00239; C2; 1.
FT NON_TER 1
SQ SEQUENCE 1053 AA; 117818 MW; D31DD27271F5E981 CRC64;

Query Match 35.4%; Score 2936.5; DB 4; Length 1053;
Best Local Similarity 54.1%; Pred. No. 2e-185;
Matches 614; Conservative 143; Mismatches 232; Indels 145; Gaps 24;

QY 1 MSAPLGRGRAPPAAPASPPPPMPDLSTTEERKTIILAVMDROKKEEKGSLTK 60
DB 26 MSSAVGGRGPPPT---VPPPMOELPDLSHLTTEERNITIAVMDROKKEEKEEEMALKC 81
QY 61 KEHKADPTQWFPFSGITELVNNVLQPOQKQNEKEPOQKHOPEMYEYOYKXKMGEEGQ 120
DB 82 VVRMAPAA-----CKTPRAENQPPQSPRLHQEFSEYEQVAKTIGEAR 128
QY 121 QQQ-EQKGDAPTCGICHTKTFADGCGHNCYCOTKFCARCGGVSLSRKNK---VAVVGN 175
DB 129 RYQGEHMDAPTCGICHTKTFADGCGHNCYCOTKFCARCGGVSLSRNNEDKVVAVVGN 188
QY 176 LCRKQOELITRSGAMFYSGSNTLQDPQKVPRLNE-----EAPQEKAKLHEQPO 228
DB 189 LCRKQOELITRSGAMFYSGSNTLQDPQKVPRLNE-----EAPQEKAKLHEQPO 242
QY 229 FQ-----GAPGLSVPAVEKGRAGHL--TRQDTIKNSGKHQIASMPDRKRSYV 279
DB 243 SQPLSTAAASQDAAPAPPPDRSKGAEPSSQALGPEQKQASRSRSPPERKKTPEGL 302
QY 280 SRDNRREYQSEEREDYSQYVPSDGTMPRSPSDYADRSQREPOFEYEPGHLNRYDSNR 339
DB 303 SEQNGKGLKERR---VPKTSAQVEGAVEERERK-----RSRL 344
QY 340 GHRKSEKIYDDEVEDSDE---YEQRRREYQYARSDPNLARYVKQYEEQKR 394
DB 345 EKGRQDY---PDPRPEKDEKKADEEKQKREDEYQTRRSPDNLARYVKKPPEEQQR 401
QY 395 IHAVSARHERRRSHSDVSLAALEDSRIILRMDRPSQORSVSEPRAMENQRSYMER 454
DB 402 MHAVSARHERRRSHSDVSLAALEDSRIILRMDRPSQORSVSEPRAMENQRSYMER 457
QY 455 TREAGQSSYQORTSNHSPPTPRSRPIPLDRPMRRADSLRKQHLDPSSAVRKTREEM 514
DB 458 TAEVRAGA--KQJLNSPPAPRHGPVAPAEALKAQELRQSLDPSSAV----- 507
QY 515 EYMLRNDLSQDSESVARPPPRPHKSKKGGKMRQVSLSSSEELASTPEYTSQDVELE 574
DB 508 -LMLRNDLSQDSESVARPPPRPHKSKKGGKMRQVSLSSSEELASTPEYTSQDVELE 566
QY 575 SESYSEKGDQKGRKRTSEQVLSNTRSEBQKRMYYGHSLEBDELMSEPOITDSV 634
DB 567 SESYSEKGDLD-----TY-----WLP----- 583
QY 635 DTCSSTLNEESHSHDKPVTWQPSKDGDRILGRIILKRLKDGSVPRDSCAGMLKVVG 694
DB 584 ---ATMHSRSTPSSHPVTWQPSKDGDRILGRIILKRT---TMRKSGALLGKVVG 636
QY 695 GKMTESGLCAFTTKVKKGLADYVGHLPQDDEVLEMMGRLLQAGATFEYVNIILSKPE 754
DB 637 GKMTDLRGLAFIRKVKKGLADYVGHLPQDDEVLEMMGRLLQAGATFEYVNIILSKPE 696
QY 755 POVELVYRPGIDIPRIPTDTHAOLESSSSFEQKMDRPSISVSPMSGRLRVQQL 814
DB 697 POVELVYRPGIDIPRIPTDTHAOLESSSSFEQKMDRPSISVSPMSGRLRVQQL 756
QY 815 SGOLSIKLMPDKVGHQILVITLIGAKDLPSREDGRPRNYPVKIYFLPDSDDKKRRRTYK 874
DB 757 PGOLSVLWMDKVGHQLIVNLQATDLPARYDGRPRNYPVKIYFLPDRDCKKRRRTYK 816
QY 875 KTLPEKWNQFIYVSHVHRERREMLTITLMDQARVEESEEGLLELTALLDEP 934
DB 817 KTLPEKWNQFIYVSHVHRERREMLTITLMDQARVEESEEGLLELTALLDEP 934

DB 817 KTLPEKWNQFIYVSHVHRERREMLTITLMDQARVEESEEGLLELTALLDEP 876
QY 935 HWYKLOTHDVSILPRPSVPLPRQLHGESPTRLQSKRLSDSEVDYDCEDGVGV 994
DB 877 HWYKLOTHDVSILPRPSVPLPRQLHGESPTRLQSKRLSDSEVDYDCEDGVGV 936
QY 995 -DYRANGRDLOSSTLSPPEQVWSSNNHCSPSPHARVIVIRTRSMSPADPPQNRVQ 1052
DB 937 PVGRSSARSKSTLTVPEQQRTHHRSRVSPIHGRMGDQKPRRLNVP-PLQKSLDEI 995
QY 1053 H--RGTATGHTYNTISRM--DRHRVMDHYSSDRDRCDAADROPYHRNSTE 1101
DB 996 HPTRRSRSPRRNHDSRSPVDHRTRDVDSQYLSQDSEELMLPRA--KRGRSAE 1047
RESULT 14
ID Q9HBA5 PRELIMINARY; PRT; 740 AA.
AC Q9HBA5;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, last sequence update)
DT 01-MAR-2001 (TEMBLrel. 16, last annotation update)
DE RAB3 INTERACTING PROTEIN VARIANT 2 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Coppola T., Magnin-Luthi S., Perret-Menoud V., Gattesco S.,
RA Regazzi R.;
RT "Molecular characterization of six different isoforms of RIM (Rab3
RT Interacting Molecule) from human brain.";
RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF263306; AAG23163.1; -;
FT NON_TER 1
SQ SEQUENCE 740 AA; 83352 MW; 09CFB918F8FE8063 CRC64;

Query Match 29.38; Score 2424.5; DB 4; Length 740;
Best Local Similarity 64.9%; Pred. No. 7.9e-152;
Matches 493; Conservative 105; Mismatches 125; Indels 37; Gaps 12;

QY 848 RPRNPYKIYFLPRPSKKNRKTIVKTLPEKWNQFIYVSHVHRERREMLTITLMDQ 907
DB 1 RPRNPYKIYFLPRPSKKNRKTIVKTLPEKWNQFIYVSHVHRERREMLTITLMDQ 60
QY 908 ARVREESEFLGELLILETALLDDEPHWYKLOTHDVSILPUPRSPPLPRQLHGESPT 967
DB 61 PVQEESEFLGELLILETALLDDEPHWYKLOTHDVSILPUPRSPPRRHINGESS 120
QY 968 RRLQSKRISDSEYSDYDCEDGVGV--DYRANGRDLOSSTLSPPEQVWSSNNHCSPGS 1025
DB 121 KKLQRSQRIQSDSDISYDEVDDIGVPPVGVGRSSARESKSTLTVPEQQRTHHRSRVS 180
QY 1026 PHRDVYQTRSMSPAPRPOQRNVEQGH--RGTATGHTYNTISRM--DRHRVMDHYSS 1080
DB 181 PHKNGDQKPRSRPLNVP-PLQKSLDEIHPTRRSRSPRRNHDSRSPVDHRTRDVDSQYLS 239
QY 1081 DRDRCDAADROPYHRNRS-----RSTEQRPRLERTTRSSRSEDPDNLMRSPSLMTGRS 1135
DB 240 EQDSEELMLPRAKGRSAECLHTSELQPLFLDRARSASTNCLRDITSL----- 287
QY 1136 APPSPALSRSHPTGVSQVTSBSPSTPGTGRGROLPLPCK-GTLERSAMDIEBRNRQKL 1194
DB 288 --HSPERERMRHQRSPQSPADTSFSSRGRQLPQVYVRSGLSIOASLIVVEERTQKMK 345
QY 1195 --NKYKQV--AGSPRLPEQYHYSKYRSGWMPHRRGADIVSTKSSQSDVSAVSTSSAS 1250
DB 346 KVHRFKQTTGSSQGLDREQYSKYNIHKDQYRSCDNDVASAKSDVDVSAISTSSAS 405
QY 1251 RFSSTSYMSVOSEPRRNRKISVFTSKMONRQMGVSKNLTKSTISGDMCSLEKNDGQ 1310

```

Db 406 RLSTSFMSGEQSERPRG--RISSEFTPKMOGRMGTSGRSLMKSTSVSGEMYTLEHNDGSO 463
1311 SPTAVGALCTSGKKRRSSIGAKMVAIVGLSKRSASQSLQTEGGKKLSTVORSTETG 1370
464 SPTAVGTGAGGKKRRSSISAKVVAIV--SRSRSTSQLQTESGHHKXSTIORSTETG 521
1371 LAVERNMWTROASRSTSGSMNSYSEGNLIPPGVRLASDSQSPDLGIGPAQLVGRQ 1430
522 MAAEHRK -MYROPRESSTGINSYSSEGNLIPPGVRLGADSQSPDLGIGPAQLVGRQ 580
1431 TLATPAMGDIOVGMDDKKQGLEVEIIRARGLYVKGSKTLPAVYKYLNDNGVCIARKK 1490
581 TLATPAMGDIOIGMEDKKQGLEVEVIRARSLTQKPSKSTPAPYKYYLLENGACIAKKK 640
1491 TKVAKTLEPLVQQLSFESPOGRVLOIIVMGDYGMDHKSFGVAQIILDELELSMV 1550
641 TRIARKTLDPLVQOSLVFESPOGKVLQVIWGDYGMHDKCFMGVAQIILLELDLSMV 700
1551 IGMFKLFPSSLVDPSTAPLTFRRASQSSLESSTGSPSYSRS 1590
701 IGMFKLFPSSLVDPSTLPLTFRRASQSSLESSTGSPCIRS 740

```

```

RESULT 15
Q9HBA6 PRELIMINARY: PRT: 766 AA.
AC Q9HBA6:
BT 01-MAR-2001 (TREMBLrel. 16, Created)
BT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
BT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE RA3 INTERACTING PROTEIN VARIANT 1 (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN:
RA Coppola T., Magnin-Luthi S., Perret-Menoud V., Gattesco S.,
Regazzi R.;
RT "Molecular characterization of six different isoforms of RIM (Rab3
interacting molecule) from human brain.";
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF263305; AAG23162.1; -.
FT NON TER 1
SQ SEQUENCE 766 AA; 86537 MW; CF492313B221C3D0 CRC64;

```

Query Match 29.1%; Score 2410.5; DB 4; Length 766;
 Best Local Similarity 63.0%; Pred. No. 7e-151;
 Matches 495; Conservative 105; Mismatches 123; Indels 63; Gaps 14;

```

QY 848 RPRNPYKLYFLPDRSDKKRTKTVKTLLEPKWQTFIYSPVHRREFRRLMLEITLMDQ 907
1 RPRNPYKMYFLPDRSDKSKRRTKTLLEPKWQTFIYSHVHRDFRRLMLEITLMDQ 60
QY 908 ARVREESFLEGLLETLALDDPRHWKLTQTHVSSLPLRPSPLYLRRLDGHGSP 967
61 PRVQEESEFLEGLLETLALDDPRHWKLTQTHDSSLPLPQSPFMPRRNTHIGSSSS 120
968 RRLQSKRISDESDVSDCEDEGVVVS--DYRNGRDLOSSTLVPQVMSNNHCSPSGS 1025
121 KKLQSRQISDSDISDIYVDGIGVPPVGYRKSARSKSTTLTLVPEOQRTTHHRSRVS 180
QY 1026 PARVDVIGTRSMSPAPPQRNVQGH--RGTRATGHVNTISRM--DRHRVMDHYSS 1080
181 PHRGNDGKPRSRRLPNV-PLQSRSLDEIHPTRRSRPTRNHDASRSPVDHRTROVDQYLS 239
QY 1081 DRD-----RDCA-----ADROPYHRSK-----TEORPLLERT 1109
240 EODSELMLM.PRAKRGSRASCLHTTTRHLVHRHYKTL.PPKMPLLOSSSHWNIYSELOPFLDRA 299

```

```

QY 1110 TTRSRSEPRDTNLMRSMPSLMTGRASPSPALSRSHPTGVSOTSPSSPTGTGRGRL 1169
300 RSASTNCLAPDTSL-----HSPERERMRHROSPTOSPPADTSFSSRRGROL 345
QY 1170 POLPK-GLTLERSAMDIEERNQML--NKYKOV--AGSDPRLQDYSKYSRSGMDPRG 1224
346 PQVPVRSQSIQASLVYEERTROMKMHVHRFKQTTGSGSSQELDRQYSKYHNKHQYRS 405
QY 1225 ADYVTRKSSDSVDVSAVSRSTSSARFSSTSYMSVQSERPRGRNKTISVFTSKNQNRMG 1284
406 CUNVASKSSDSVDVSAISRTSSASRLSTSEMSQSERPRG--RLSSFTPKMQGRMG 463
QY 1285 VSGKNLTKSTISGDMCSLEKNDGSDQTAVALGTGKKRRSSIGAKMVAIVGLSKRSR 1344
464 TSGRSTMKSTSVSGEMYTLEHNDGSOQDVAIVGYAGAGKKRRSSLAKVVAIV--SRSR 521
QY 1345 SASQSLQTEGGKKLRSTVORSTETGLAVERNMWTROASRSTSGSMNSYSEGNLIFP 1404
522 STQSLQTESGHHKXSTIORSTETGMAAEHRK -MYROPRESSTGINSYSSEGNLIFP 580
QY 1405 GYRLASDSQSPDLGIGPAQLVGRQTLATPAMGDIOVGMDDKKQGLEVEIIRARGLYVK 1464
581 GYRLGADSQSPDLGIGPAQLVGRQTLATPAMGDIOIGMEDKKQGLEVEVIRARSLTQK 640
QY 1465 PGSKTLPAPYKYYLLNDNGVCIARKKTVAKTLEPLVQQLSFESPOGRVLOIIVMGD 1524
641 PGSKSTPAPYKYYLLENGACIAKKKTRIRKTLDPVQOSLVVDESPOGKVLQVIWGD 700
QY 1525 YGRMDHKSFGVAQIILDELELSNMVYIGWFKLPPSSLVDPSTAPLTFRRASQSSLESSTG 1584
701 YGRMDHKSFGVAQIILLELDLSMVYIGWFKLPPSSLVDPSTLPLTFRRASQSSLESSTG 760
QY 1585 PSYSRS 1590
761 PPCIRS 766

```

Search completed: November 21, 2001, 16:09:59
 Job time: 265 sec

Fri Nov 23 10:08:44 2001

us-09-617-099b-2.rge

FD/PO 7/14 (2000)
Rev. 2 & Rev (JBC) Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2001, 15:57:54 ; Search time 8528.31 Seconds
(without alignments)
9032.199 Million cell updates/sec

Title: US-09-617-099b-2
Perfect score: 4980
Sequence: 1 gcttcctaggggtgttcgg.....acattgtgccttagcaaaag 4980

Scoring table: IDENTITY_NMC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rdg:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rdg:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vl1:*
59: gb_vl2:*
60: gb_hcg1:*
61: gb_hcg2:*
62: gb_hcg3:*
63: gb_hcg4:*
64: gb_hcg5:*
65: gb_hcg6:*
66: gb_hcg7:*
67: gb_hcg8:*
68: gb_hcg9:*
69: gb_hcg10:*
70: gb_hcg11:*
71: gb_hcg12:*
72: gb_hcg13:*
73: gb_hcg14:*
74: gb_hcg15:*
75: gb_hcg16:*
76: gb_hcg17:*
77: gb_hcg18:*
78: gb_hcg19:*
79: gb_hcg20:*
80: gb_hcg21:*
81: gb_hcg22:*
82: gb_hcg23:*
83: gb_hcg24:*
84: gb_hcg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------------|--------------------|
| 1 | 4608.4 | 92.5 | 4800 | 94 AB021131 | AB021131 Mus muscu |
| 2 | 3886 | 78.0 | 5550 | 94 AF199331 | AF199331 Rattus no |
| 3 | 3834 | 77.0 | 5592 | 94 AF199330 | AF199330 Rattus no |
| 4 | 3776 | 75.8 | 5640 | 94 AF199322 | AF199322 Rattus no |
| 5 | 3197.6 | 64.2 | 5172 | 94 AF199324 | AF199324 Rattus no |
| 6 | 2936.4 | 59.0 | 4965 | 94 AF199329 | AF199329 Rattus no |
| 7 | 2860.4 | 57.4 | 5031 | 94 AF199323 | AF199323 Rattus no |
| 8 | 2860.4 | 57.4 | 5031 | 94 AF199326 | AF199326 Rattus no |

|||||
Db 421 GACCTGTGGCATCTGCACAGACAAATTTTGAGATGATGCGGCATATTTGTTCTTA 480
Qy 481 ttgccaacacaaagttctgtgtcgcgatgtgaggtcgaagtgtcttcaagtcacaagagt 540
Db 481 ttgccaacacaaagttctgtgtcgcgatgtgaggtcgaagtgtcttcaagtcacaagagt 540
Qy 541 tatgtgggtgtgtaattttgtgcgaaacacaaagaatctctcaataatcagagatg 600
Db 541 TATGTGGGTGTCTAATTTGTGCCGAAAAACACAGAAATCTCTACTTAATCAGAGATG 600
Qy 601 gtttataatagtggtgtctcaacacactgcagcaacctgtalcacaaggttctctcgaaggt 660
Db 601 GTTTATATATAGTGGGTCTAACACACTGCAGCAACTGATCAAAAGGTTCTCTGAGGGCT 660
Qy 661 tcgaaatgaagagacccctcgaagagaagaagaacatacagaagcagcccaagttcca 720
Db 661 TCGAAATGAGAGAGCCCTCCAGAGAGAAAGCAAACTACACGAGCAGCCCAAGTTCCA 720
Qy 721 aggaagcccaagtgactta tcagtaacctgcagtgtgaagaagcgcgagctatgggtcac 780
Db 721 AGGAGCCCAAGTGACTTATCTAGTACTCTGCAGTTGAGAAAGGCCGAGCTCATGGGCTCAC 780
Qy 781 aagacaggaatactatlaaanaatggaatcgaagagtgaaagcaccagattgccaagtgcac 840
Db 781 AAGACAGGATACTATTTAAATGAGATCAGAGTGAAGACACACAGATTGCCAGTGACATGCC 840
Qy 841 ttcagaagaanaacgaagtccatcagttgccaaggtatcaaaatcgaaagatacagaagaag 900
Db 841 TTCAACACGAAAAACCAAGCCATCATGATGTCAGGGATCAAAATCGAAGATCGACGCAAG 900
Qy 901 tgaagaagaagaagaactctacacagta tgccttcagatgtgtaacatgtcagaagatctcc 960
Db 901 TGAAGAAAGAGAGACTACTCTACAGTATTTCTTTCAGATGATGATCAATGCGCAAGATCTCC 960
Qy 961 ttcggaattatgctga tagaagca tctcagcgttgagccctcaatttatagaagaacctgtlca 1020
Db 961 TTCGATATATGCTGATGAGAGCATCTCAGCTGAGCCTCAATTTTATGAAACACCTGTGTC 1020
Qy 1021 tttaaattacaaaggtatcttaacagaagaagccatagacatctccaagagtatatgtgga 1080
Db 1021 TTTAAATTTACAGGGATTTCTTACACAGGAGGCCATGACATTTCCAAAGATATATTTGGA 1080
Qy 1081 tgaataagaatgtggagagaagaatgaatataaagaacaaaggaagaagaagaatcca 1140
Db 1081 TGATTAAGATGTGGAGAGCAGATGAATATGAAAGCAAAAGAGAGAGGATACCA 1140
Qy 1141 ggcagctacagaagtgatcccaaatctgccccggtatcccgtaagccacaacctcaga 1200
Db 1141 GGCACGCTACAGAAATGATCCAAATCTGGCCCGGTATCCCGTAAAGCCACAACCTTACGA 1200
Qy 1201 agaacaatgcatccatccagctggaaggtgtccaaggcagaacatgagaagaagcaagtga 1260
Db 1201 AGAACAAATGCGCATTCACAGCTGACGCTGTCCAGGGCACGACATGAGAGAGCACAGTGA 1260
Qy 1261 tgtttctttggaagaagcgtgaactagaagaatcccaagatttctctgtctaagatgtgtag 1320
Db 1261 TGTTCCTTTGGCAAGCGCTGAACTAGAAATTCACAGATTTCTCTCTTAAGATGGATG 1320
Qy 1321 aacatcaaggcaagaatctgtatctgaaagtagaagctgcaatggaanaacaaagatcgtta 1380
Db 1321 AACATCAAGGCAAAATATCTGTATCTGAAGCTGAGCTGCAATGGAANAACCAACGATGTGA 1380
Qy 1381 ttcaatggaagaactcgaagaggtcgaaggacaagttcttatccacaagaagaccacca 1440
Db 1381 TTTCAATGGAAGAAGAACTCTGAGAGGCTTCAGGACAAAGTTCTTATCCAAAGGAGACTTAAA 1440
Qy 1441 tcaatagctctccaccctccgagcgagccctatacgcgtgataagaccacaagatgaagg 1500
Db 1441 TCATAGTCTCTCCACCCCTCGCGGAGCCCTATACGCTTGATAGACACATGAGAGCG 1500
Qy 1501 cgctactccctactggaacacagacacacttaagctccagctcgtcgtgtggaagaagaa 1560
|||||

Db 1501 CGCTACTTCCTACGGAACAGCACCACTTAGATCCAGCTCTGCTGTGAGAAACGAA 1560
Qy 1561 gcgagaaaaaatgnaaacatgtltaaggaaatgtatctttagtltcagaacgaagtcgggtc 1620
Db 1561 GCGAGAAAAAATGAAACCATGTGTAAGAAATGATTTCTTTAGTTCCAGACAGTCCGAGTC 1620
Qy 1621 agtgaagccgcccccaagcgcataaatcaagaagaaggtlaaagtgcagcaggt 1680
Db 1621 AGTGAAGCCGCCCCACCAAGGCTCATTAATCCAGAAAGAGNGTAAATGCCCCAGGT 1680
Qy 1681 ttcaatgaacagctcgaagagaagctggaatccacaacctgaaglatacaagctgtga tga 1740
Db 1681 TTCACGTGACACCTCTCGAGAGAGAGCTGGCATCCACACCTGAGATATCAAGCTGTGATGA 1740
Qy 1741 tgtgagctcgtgaagaagcgaagtgtgagtgagaagaagggaacgtccaagaaggaaaaaaa 1800
Db 1741 TGTGAGCTGTGAAAGCGACAGTGTGAGTGAAGAAAGGAGACATCAAAAGGAAAAAAGAAA 1800
Qy 1801 aactagtagcagggaggtttgtgcgattc taaacacaggtctgagagacaaaagaag 1860
Db 1801 AACTAGTAGCAGGAGGATTTGTGCGATTTCTAACACAGGCTGTGAGAGACAAAGAAAG 1860
Qy 1861 gatgtactatggtgacctctttggaagaagga tttggaatggtctgagcctcagatlaa 1920
Db 1861 GATGTACTATGTGTGGCCACTCTTTGGAAGAGAGATTGGAATGCTGTGAGCCTCAGATTAA 1920
Qy 1921 ggaactctgggttagataactgtgtagcacaaaccttaagaagaagcatgccaatga tga 1980
Db 1921 GGACTCTGGGGTAGATACCTGTATGATGACACACCTTTAAAGAGAGATGAGCCCTAATGTA 1980
Qy 1981 taagcaacctgtgaacctgtgcagccatccaaaga tggagatcgccataattgtgtatltt 2040
Db 1981 TAAAGCACCTGTGACCTGGCAGCATCCAAAGATGAGAGATCGCCTAATTTGCTATATTTT 2040
Qy 2041 altaaataegcgtltaanaagatggagtgatccctcgagactcagaagcaatgctgggct 2100
Db 2041 ATTAATAATAGCGTTTAAAGATGGAGTGTACCTCGAGACTCAGAGACCAATGCTGGGCTT 2100
Qy 2101 aaaggttctgaagagaagaatgagctgaatcaggtcgagcttctgtcatlttatccaaagt 2160
Db 2101 AAAGGTTGTAGAGAGAAATGACTGAATCAGGTGCACTTTGTGCAATTTATACCAAGT 2160
Qy 2161 aaaaaaagaagtttaagcgtatactgttagacatctttagacaaaggtatgaatctctgga 2220
Db 2161 AAAAAAAGAAAGTTTATGCTGATCTCTAGGACATCTTTAGCCAGGTGATGAGTCTTGGA 2220
Qy 2221 atggaatggagagctatgtgcaagaagccacattggaagaagtttacacatatctcaga 2280
Db 2221 ATGGAATGGAGAGCTATGTGCAAGGAGCCACATTTGAGGAAGTTTACAAATATTTCTAGA 2280
Qy 2281 atccaaacctgaaccacaagaattgagcttgttltcaaggccaaattggagatatctctag 2340
Db 2281 ATCCAAACCTGAAACCAAGATGTAGCTGTGTTTGAAGGCCAATGTGAATATATCTCTAG 2340
Qy 2341 aatactcgtatgaacagcagcagaactggaatccgaatctcagatcttaagatctcaaaa 2400
Db 2341 AATACTGATGACAGCAGTACACAACTGGAATCCAGTTCTAGCTCATTTGATATCTCAAA 2400
Qy 2401 aatgacagctctctctataltcgtlaccctcaaccatgagctccggtcagtgtcgaagga tgt 2460
Db 2401 AATGAGCCGCTCTTATATATTCGTTACCTTACCTCACCATGAGTCTCGCATGTGAGAGATGT 2460
Qy 2461 cccgcaagttcttatctgtgaacagcttcaataaactatggtttgacaaggttgtgtcaaca 2520
Db 2461 CCCGCAAGTCTTATCTGAGCAGCTTCAATTAACATATGTTTGCACAAAGTGTGCTCACCA 2520
Qy 2521 gttgtagtgaatatttggagcaagaagatctcccttcaagggaagaatggagggcgaag 2580
Db 2521 GTTGATATGATTAATTTTGGAGCAAAAGATCTCCCTTCAGGGAAGATGGAGGCGCTAAG 2580
Qy 2581 gaatccctatgttlaagaattactctctccaga tgaagtgataaanaataagaagaagac 2640
Db 2581 GAATCCTTATGTTAAGATTTACTTCTCCAGATAGAAAGTATAAATAAAGAAAGAAC 2640
|||||

Db 4621 ATGCACTATTAACCTTTGTCACAAACACGCGATCAAAACCGAAGAAACCGA 4680
 Oy 4661 caggtcggaagcccccgtggaacactgcacgtcctgtagctgtgtctacagaagccacgtcta 4920
 Db 4661 CAGGTGGAAGCCCCCGTGAACACTGCATGCTTGAGTGTGTGTCTACAGACCCACTGCTA 4740
 Oy 4921 gggataccaagcagtcctgtgttctcagaaggaagtcgtacacattggtccctagcaag 4980
 Db 4741 GGGATACCAAGACGCTCGTGTCTCAGAGGAAGCTGACATGTCGCCCTAGCAAGG 4800
 RESULT 2
 LOCUS AF199331 ROD 04-JUL-2000
 DEFINITION Rattus norvegicus RIm2-5B (RIm2-5B) mRNA, complete cds.
 ACCESSION AF199331
 VERSION AF199331.1 GI:8925875
 KEYWORDS
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 5550)
 REFERENCE 1
 AUTHORS Wang, Y., Sugita, S. and Sudhof, T.C.
 TITLE The RIM/NIM family of neuronal C2 domain proteins. INTERACTIONS WITH RAB3 AND A NEW CLASS OF SRC HOMOLOGOUS 3 DOMAIN PROTEINS
 JOURNAL J. Biol. Chem. 275 (26), 20033-20044 (2000)
 MEDLINE 20347919
 REFERENCES 2
 2 (bases 1 to 5550)
 AUTHORS Wang, Y. and Sudhof, T.C.
 TITLE Direct Submission
 JOURNAL Submitted (27-OCT-1999) Center for Basic Neuroscience, The University of Texas Southwestern Medical Center, 6000 Harry Hines Blvd., Dallas, TX 75235-9111, USA
 FEATURES
 SOURCE 1..5550
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 1..5550
 /gene="Rim2-5B"
 52..4629
 /gene="Rim2-5B"
 /codon_start=1
 /product="Rim2-5B"
 /protein_id="AAEP1653.1"
 /translation="MSAPLGRGRAPAPPAASQPPPEMPDLSTLEERKILIAQVAM
 DROGKEEKEDSVYAKLHQOEFEMKKEOVKKKEGSESOQOEDGAPTCIGIKHTKPAD
 GCGNGSYCOTKPCFACGGRGYSRLSNKNTVMWVCNLCKRQOELTSGAFVYSGNTPO
 OPKQALRGALRSEAPROBKAKLHQEPOFOGPPDDSSVAYERRANGILTRQDSIKNG
 SGKHQALSDMRSDRKRSPSVSRQNRNTRIDSEEREESQVPEDSSTMPRSPSYADR
 RSQEPQEEPEPHLANRDSNRGRHRSKEYIVDEVESESDVEDEKORREBEYQARY
 SDPMIAPYVKKQPEYEQEMRIHAESVARSHEHRRSDVSLAAELSDRSILSLMDRPSR
 RQRESEERRAMENORSTYSMERTREAGOSSTYPORTNHSPTPRRSIPLDRPRLRS
 ADSLRKHLDHPDSAVVTRKREKMETLRLNDSLSDSOSESPRPPPHKSKGKKMR
 QVLSRSEEBELASTPEYTSQDDVITESESGEKKDMETSNLEHNSHMSSEKSPMSLHP
 VTWQPSKDGDLIRILRLNLKKGVSYPDSGAMGLKLVGKGTESGRICAFITTKVK
 KGSLATDVGLRPDEVLNENGRLLQGAETEEVNIILSESPQEVLELVSPRGDMH
 RLPDSPTHAQLESSESSSEFESQMDRPSISVSPKSGMLRDVQFLSGLIKLWFDMF
 GHQILVILGAKDLPSREDDGRPNRYKIVYLPDRSKNRKRTYVTKLTPPKKNQNE
 IYSVHRREPERBELTETIMQOARVEESEFGEITLIELTALDQPHMYKLTQHDH
 VSSLPSPHPYMRROLHSGSPRRRQRSKRISDSEYSTDCEDGYVSDYRHDGR
 DLSSTTSPEEQVMSHNCSPSPRHVDYIGTRRSVSPVQPRNVEDLKRGTARV
 GHVYTSRMDRHRVMDHYSSERDSHLTLESRHHROTSEHHNDGRCEADARQVPH
 RSRSTQEPRLLERTTTRSRSEERADTNLMHSMPLMTGRSAPSPALSRSHPRGVSQ
 TSSSTPTVTRGRGLPOLPPKGLERSANDIIEERNOMKLNKYKQVAGSDPRLEQDY
 HSKYRQWGDHGRGADVSTKSSDSVDSVASTSASAPRSSTVMSVOSERGRGNR
 KISVTSKMSKRMQMGVSGKMAKTSISGDMCSLEKNDGSQDPAVALGTLISGKKRRS
 SIGKMWAVIGLSKRSKRSASOLSQTDEGGKRLRSTVORSTETGLAVEKRNMTQDAR
 EMDKQGSNLSYSEGNLTFPEVRLASDSQSFQDFDLGPAQLVGQTLATPMGDIOVG
 MMDKKGQLEVEIIRARGLVVKPGSKTLPAPVYVLLDNGVCIARKKTKYARKTLEPL

[illegible]

| | | | |
|----|------|--|------|
| Db | 2880 | TTTATAGGAAGACTTAGGCTCATGGCTGCGCTTAGTGTCCTCCCTCCCTCAAGAGAAATGTGGGA | 2939 |
| QY | 3181 | acaagagcacccgaaggagacacgctgctactctgacctatcaaacacatlagccgaatggaatag | 3240 |
| Db | 2940 | ACAGGGGCGCTTCAGAGGACACGCTGCTACTGCGCATTCACATACATATRACCGGATGGACAG | 2999 |
| QY | 3241 | acaaccgctcatalgatalgacactactctctcagaatagagac----- | 3280 |
| Db | 3000 | ACACCGTCTCATGACGACACCATTACTCTTCAGAGAGACAGATTCATTTCTCACTTAC | 3059 |
| QY | 3281 | -----agggattgtgaagc | 3294 |
| Db | 3060 | TCGCTCTCGACACAGGACAGACCAATGATGATCACCAACAGGGATGAAAGGGATGTGGAGC | 3119 |
| QY | 3295 | agcaatatagacagccataltcacagatccagatccaagaacaacgacctctccatagaagc | 3354 |
| Db | 3120 | AGCAGATGACAGACCCATATCACAGATCCAGATACACAGAACCAACGGCCTCTCCTAGAGCG | 3179 |
| QY | 3355 | gaccacaccccgctccagatlcctcttgaaagctcctgatalacaacctcgaatgaatgcatgccc | 3414 |
| Db | 3180 | GACCAACACCCGCTCCAGATCCTCTGAACGTGCTGATACAAACCTCATGAGGTGATGCC | 3239 |
| QY | 3415 | ttcataatgaactggaaagatctgccccctctcaacttgctcttctgaagtctcaacctcg | 3474 |
| Db | 3240 | TTTCATTAATGACTGGAAAGTCTGCCCCCTCTTCACCTGCTTATCGAGGTCTCACCTCG | 3299 |
| QY | 3475 | taccgggtctgtccagagacaagcccatcaagtctccgggagacaggaagggccgacaca | 3534 |
| Db | 3300 | CACGTGGCTGTGTCACAGACAGCTCCGTCGAAGTACTCCCGTATACAGAGCAAGGGGCGCGCA | 3359 |
| QY | 3535 | gcttccacaagcttccacaccaaaggaaacatlvgagaaagtgctatgatalatagagagag | 3594 |
| Db | 3360 | GCTTCTCTCAAGTCCCAACCAAGGAAGACTGGAAGAAGCTATGATATATAGAGAGAG | 3419 |
| QY | 3555 | aaatctgccaaatgaaacttaacaataacaacaaaggtagccgagatcagaagccccaacttga | 3654 |
| Db | 3420 | AAATGGCCAAATGAAACCTTATACAAATACAAACAGGTATGCGGATCAATATCCCAACATGGA | 3479 |
| QY | 3655 | gcaagattacacatlcgaaglatcgcctccagatavggatlcacatagaaagggcagaatctgt | 3714 |
| Db | 3480 | ACAAGATTAACCTTGTGAAGTATTCGCTCAGAGATGGGATCCACATATAGAGGGGCGAGTACTGT | 3539 |
| QY | 3715 | ttccaataatccctcggacaagtatglaagtgaatglatcctgcyggttccaaggaatagtag | 3774 |
| Db | 3540 | TTCTAATTAATCTCGGACAGATGATGATGATGATGATGATGCGGTTCAAGAGACTTACTAG | 3599 |
| QY | 3775 | tgcttctcglttcagcagcacaaagctacatgltccgtccaatcagaagcgcgcgaagagaa | 3834 |
| Db | 3600 | TGCTTCTGCTTTACAGACAGCAACAGTACATGTCCTGTCATTCAGAAAGCGCCGAGAGAAA | 3659 |
| QY | 3835 | caggagaataatcagatgctttaccatccaaaatgcaaaaacagacaatlggctgtgtgggaa | 3894 |
| Db | 3660 | CAGGAAATACATGTGCTTTATACGTCCAAATATGCAAAACAGCAACAAATGGGTGTGTGGGGAA | 3719 |
| QY | 3895 | gaacttgaaccaaagaacaccagacatacagttgagacatgctcactgtagaagaatgtagcgy | 3954 |
| Db | 3720 | GAGCATGTGGCCAAAGACACGATCATGCTGTGAGACATGTGCTCACTGGAGAAGATGATGG | 3779 |
| QY | 3955 | cagccagtcgcgaactgcaagtvggcygcctgggtaccagatgycgaagaagcgcgcgaatcag | 4014 |
| Db | 3780 | CAGCCAGTCTGACACTGCGAGTGGGGCCCTCGGCGACCAAGTGGCAAGAAAGCGCGCATTTAG | 3839 |
| QY | 4015 | catgtggggccaaaatggtatgctatgtgtgtctcctaagaaaagtcgcaagtgtctctca | 4074 |
| Db | 3840 | CATTGGGGCCAAATATGATGCTATATTTGTGGTCTCTCACGGAATAATGTGCAATGCTTCCA | 3899 |
| QY | 4075 | actcagaccaaaacgaagagagatgataaaagctacagagagcaactgttccagaagaacacagga | 4134 |
| Db | 3900 | ACTCAGCCAAACCGAAGGAGAGGTAAAAAGCTATCCGACGACTGTCCAGAGAAACACAGGA | 3959 |
| QY | 4135 | gaccgggtacgagtggagatgtaggaactgtgatacccgccagggccagccggygaatccac | 4194 |

|||||
Db 1221 ACCATCAAGCAAGATCTGTATCTAGCTAGCTGCAATGAAACAGCATGCTA 1280
Oy 1381 tccaatggaagaatctcgagagctcagaggacaagaattctatccacaaggacctcaaa 1440
Db 1281 TTCAATGGAAGAACTCGGAAGCTGAGGACAAAGTTCTTATCCACAAAGGACCAAAA 1340
Oy 1441 tcaatgctcccaaccccccgagccctatacgccttgatagacagacatgagcg 1500
Db 1341 TCATGCTCTCTTACCCCTTCGACAGAGCCCTATACCGCTGATRAGACCAAGCTGAGGCG 1400
Oy 1501 cgcctgctccctcaggaagaacagcaccaactagaatcccaagctctgctgtaggaaacgaa 1560
Db 1401 TCCGACCTCCCTACGGAACACACACACCACTTAGATCCAGCTGCTGTATGAGAAACGAA 1460
Oy 1561 gcgagaaaaaatggaacaacatgtaagaatgattctcttgatgcagacagtc 1620
Db 1461 GCGAAGAAAAATGGAACAACTGTGAGGAAATGATCTTTGAGTTCAACACAGCTCGAGTC 1520
Oy 1621 agtgaagccgcccccaagaagccataaataccaagaagaaggtataatgacaggt 1680
Db 1521 AGTGAAGCCGCCCCCAACAGCCCTCTAATATCCAAAGAGAGTAAATGCCCAAGCT 1580
Oy 1681 ttcactgaagcaagctcgaagagagagcttgacatccacacactgaatatacaagctgtga 1740
Db 1581 TTCACCTGACAGCTCAGAGAGAGAGGATTGGCATCCAGCCTGAGTATACAAAGCTGTGATGA 1640
Oy 1741 tctgaagcgcgaagaagcgaagtgtagtgtagaagaaggagacgtcaaaaggaaaaaaga 1800
Db 1641 CCGTGGAGATTGGAAGCGAAGCGTAGCTAGGAAGAGGAA----- 1681
Oy 1801 aaactagtgacgaagagttgttcgagatcttaacacacaggtctgagagacaaagaaga 1860
Db 1682 ----- 1681
Oy 1861 gatgactatggttgccactctcttgaaagaggaattggaatggtctgagctcagatcaa 1920
Db 1682 -----TGAGTACAGCTGGTTGAGCATGCGCTCTTG 1712
Oy 1921 ggaactcgggtagataactgttagtagcacacaaccttaacgaagagacatagcagtagtga 1980
Db 1713 G-----CATAGCAGTGAGGACCAATCCCAATGTC 1739
Oy 1981 taagcacctgtgacctgacgcagccatccaaagaatgagatcgccctaaattggtgatttt 2040
Db 1740 TTTCGACCTGTGACCTGGACGCATCCAAAGATGAGATCGCTTAATGGTCTGATATTT 1799
Oy 2041 ataaataagcgtltaaaagatlgagagtgctaccccgagactcagagcaatlgctggcct 2100
Db 1800 ATTAATAATAGCGTTTAAAGATGAGTGTGACTCGAGATTCAAGAGCAATGCTGGGCTT 1859
Oy 2101 aaaggtctgaggaagaagaatgactgaatcaggtcgacttgagcttattatccaaagt 2160
Db 1860 AAAGGTTGAGGAGGAAATGACTGATCATGAGTGCAGCTTTGTGCAATTTATTACCAAGT 1919
Oy 2161 aaaaaaagaagaatttagctgatactgtagagacatcttaagccaggtgataaagcttggga 2220
Db 1920 GAAAAAAGGAAGTTTAGCTGATACGTGAGACATCTTAGACCAAGATGAGAGCTTGGA 1979
Oy 2221 atggaatggagagctatctgcaagagacacattggaagaattacaacatlatctaga 2280
Db 1980 ATGGAATGGAAGGCTATTGCAAGAGACACATTGAGGAAGTTTACAACATTTATTTAGA 2039
Oy 2281 atccaaactggaaccaagaagttagactgtgttctcaaggccaattgagatatctctag 2340
Db 2040 ATCCAGAGCTGGAACCAAGATTGAGCTTGTGTTTCAAGGCCGATTTGAGATATGCTAG 2099
Oy 2341 aatcactgtagcagcagacagcaactggaatcaggtctagctcatattgaatccaaa 2400
Db 2100 AATTACTGTATGACTCATGACACAACTGGAATCCAGTTCTTAGCTCATTTGATGTCAAAA 2159
Oy 2401 aatgagccgtctctatatactcglttaacctaacccatgagtcgagcagtcgaggaagt 2460
|||||

Db 2160 AATGATCGTCCGCTATATCGTTACTCTCCCATGAGTCTGCGATGCTGAGGAGT 2219
Oy 2461 cccgcagttctatctbgacagcttcaataaactatggttgaaagttgtaacca 2520
Db 2220 CCCACAGTTCTTATCTGGACAGCTTTCAAATAAATGATGTTGACAAAGGTTGCTCACCA 2279
Oy 2521 gttgaagtataaatlttbgagcaagaagatctccctccaaggaagatgagagccaag 2580
Db 2280 GTTATATGATTCAATTTTGGACACAAAGATCTCCCTCCAGGGAAAGATGGAGGCCAAG 2339
Oy 2581 gaatcctatgttaagaattacttccctccagaatagaatgataaataagaagaagac 2640
Db 2340 GAATCTTATGTTTAAATTTTACTTCTTCACAGACAAGATGATTAATAAACAAGAAAGAAC 2399
Oy 2641 aaaaacagtcgaagaactlttgaaacccaatggaaccagacttcaattatctcctgt 2700
Db 2400 AAAAACAAGTCAAGAAAACTTTGAAACCCAAATGGAACCAACATTTCAATTTATCTCTGT 2459
Oy 2701 ccaccgaagagaatcccgtaacgaaatgctggaatlaacctlttgatcaagctagaat 2760
Db 2460 TCACCGAAGAGATTCGGGGAACGAAATGCTCGAATCCACCCCTTTGGATACAGCTGAGT 2519
Oy 2761 tcgagaagaagaagcgaaatctccttaagagagaatttaattgaatlgaaacagcttgct 2820
Db 2520 TCGAGAGGAAGAAAGATTCCTTAGAGAGATTTTAAATGAAATTTGAAACAGCTGTGT 2579
Oy 2821 agatagtagccgcactgtgtataagctgcagaccatgtagtctctcatctgcacctccc 2880
Db 2580 AATATATGAACACACTGTGTACAACTTCAGACCAATGATGTCTCTCATTTGCCACTCC 2639
Oy 2881 tgcacctcccatatctccccgagcgagctccatlgagagagcccaagcgcaagct 2940
Db 2640 TCACCTTCTCATATATTCGACAGAGACAGCTGCAATGAGAGAGCCACACAGCAAGTT 2639
Oy 2941 gcaaaagtctgaagaagaataagtgacagtgtaagtgtctgactacgactcgagatggcgt 3000
Db 2700 GCAAAAGGTGGAAGAGAAATGATGACAGCAAGATGATCGAATGATGACGAGGCGGCT 2759
Oy 3001 gggagtgaagtcaagatatacgcacacaatggccgagctcttcaaaagctccagctgtgct 3060
Db 2760 GGGAGTGTGTAGATTACCGACACGATGGCCGTGATCTTCAAGCTCCACATTTATAGT 2819
Oy 3061 gccagaacaagtcaltgcatcaaatcatgtctcacaatcgaaggtctcctcatcgagtaga 3120
Db 2820 GCCAGAACAGTCAATCTCATCAAAACCATTCCTCACCATCGCTCTCTATGAGTAGA 2879
Oy 3121 tgtataagaagacaaggtcaltgctgccttagtgcctcctcctcctcaaaagaatlgga 3180
Db 2880 TGTATAGGAAGACTAAGTTCATGTGCGCTAGTGTCTCTCCCTCAAAAGAAATGAGA 2939
Oy 3181 acaagggcacccgaaggacacgtgtactctggccattacaacaaattagccgaatgtagtag 3240
Db 2940 ACAGGGGCTTCAGAGGACACGTGCTACTGGCATTTCAAAATCAATTAAGCCGAGATGACAG 2999
Oy 3241 aacacgtgtcaltgagtagtgaactactccttaagaatagaagc----- 3280
Db 3000 ACACCGGTGATGACGACGACCATTTACTCTTCAGAGAGACAGTCATTTTCTCACTACCC 3039
Oy 3281 -----agggaattggaagc 3294
Db 3060 TCGCTCTGACACAGGACAGACAGTAGATCACCACAGAGGATGGAAGGATTTGTGAAGC 3119
Oy 3295 agcagatagaacagccatatacagaatccagaatcaacagaacacagcctctcctaaagcg 3354
Db 3120 ACACATATGACAGCCCTATACAGATCCAGATCAACAGAAACAGGCGCTCTCCATGAGCG 3179
Oy 3355 gacacacacccgtccagatctcctctgaacgctcgtgataaacaactcaatggatgtagtcc 3414
Db 3180 GACCAACACCCGCTCCAGATCTCTGACAGCTGCTGATACAAACCTCTATGAGGTGATGCC 3239
Oy 3415 ttcattaatgactggaagaatctgcccctccttcaactgacctatcgaggtctcaacctcg 3474
Db 3240 TTCAATTAATGACTGGAAGATCTGCCCTCTTCACCTGCTTATCGAGGTCTACCCCTCG 3299
|||||

Db 1581 TTCACCTGAGCAGCTCAGAGAGAGATTGGCATCCAGCCTGAGTATACAGCTGTGATGA 1640
QY 1741 tctggagctcggaaagcgaagtgctgagtgaaagaagggacgctcaaaagggaaaaaaga 1800
Db 1641 CGTGGAAATTGGAAAGCCAGACGCTAGGTGAGAAAGAGAGACAT----- 1682
QY 1801 aactagtgagcagggaggttctgcgattctcaacaccagctcgtgagagacaaagaag 1860
Db 1683 ----- 1682
QY 1861 gatctactatgctggtccaccctcttggaagaggalttgaaatggtctcgtccacgattaa 1920
Db 1683 -----GGAGTACAGCTGTGGTGGACATGCGCTCTTG 1712
QY 1921 ggaactcgggtagatacctcgttagtaacaaaccttaacgagagagcctatagccatagta 1980
Db 1713 G-----CATAGCAGTGGAGCATCCCAATGTC 1739
QY 1981 taagcacacctgtgacctgagccatccaaagatggagatcgcccaattggtcgtatttt 2040
Db 1740 TTTGCACCCCTGTGACCTGGCAGCATCCAAAGATGGAGATGCCCAATTGGCTGATTTT 1799
QY 2041 atlaaataagcgtttaaagaatgggagtgtaacctcogagactcaggaagaaatgctgagcct 2100
Db 1800 ATTAAATTAAGCGTTTAAAGATGGAGATGTAAGTCTGAGATTACAGGACCAATGCTGGCTT 1859
QY 2101 aaaggttgtgagggaaaagatgactgaatacagtcgacttggcatttattaccaaagt 2160
Db 1860 AAGGTTGTAGAGAGAAAGATGACTGAATGAGTGAGCTTGTGCTATTTATTAACCAAGT 1919
QY 2161 aaaaaaagaagtttagctgtaactgtgtagaacaatcagcaagtgatgaagtcctgga 2220
Db 1920 GAAAAAGGAAGTTTAGCTGATACTGTAGGACATCTTAGACAGGTGATGAAGTGTGGA 1979
QY 2221 atggaaatggagagctattcgaaagagcccaatttggaggaaatttaacaatlatctaga 2280
Db 1980 ATGGAATGGAAGGCTATTGGAGAGGCCAATTGGAGAACTTTAACACATTATCTCTGA 2039
QY 2281 atccaaaccggaaccaagaattgagctgtgtgttccaaggccaattggagaatattccctag 2340
Db 2040 ATCCAAAGCCCGAAGCCACAAATTTGAGCTTGTGTTCAAGGCCGATTGGAGATATGCCCTAG 2099
QY 2341 aataactgtagaagaatgacaaatgcaactggaatccagttctagctcaatttgatctcaaaa 2400
Db 2100 AATACCTGATAGACATCATGACAACTGGAATCCAGTTCTGATCATTTGATTCATAAA 2159
QY 2401 aatggacgcttctctatataccgcttacctcaaccaatgaatccctgagcgtcgaggatgct 2460
Db 2160 AATGATCGTCCGCTATATCGGTTACCTCCCATGATGATGCTGGCATGCGAGGATGT 2219
QY 2461 ccgcagcttcttactcggagacgttca----- 2488
Db 2220 CCCACAGTCTTATCTGAGACAGCTTTCAAGCCAAAGCCTTAGTGAAGAACACAGCCCTTT 2279
QY 2489 -----ataaactatggtttgacaaggttggttaccaggttgatagttac 2532
Db 2280 TGTTCCTAGAGGTTTCAGATTAATACTATGTTGACAAAGTTGGTTCACACGATTAATAGTTAC 2339
QY 2533 aatttgggaagaagaatccctccacagggaagaatggagagccaaaggaatccctatgct 2592
Db 2340 AATTGTTGGAGCAAAAGATCTCCCTTCCAGGAAAGATGGGGGCGCAAGGAATCTCTTAGT 2399
QY 2593 taagatttacttctccagatagaatgtatataaataagaagaacaataaagttcaa 2652
Db 2400 TAAATTTTACTTCTTCCACACAGAAATGATTAATAACAAAGAGAAACAAAACAGTCAA 2459
QY 2653 gaaaacttggaaacccaatggaaaccagacttcaattatctcctgtccacgaaagga 2712
Db 2460 GAAACTTTGGAAACCAATAATGGAACACAGACTTTCATTTATTTCTCTGTTCCACCAAGAGA 2519
QY 2713 attcogtgaagaaatgctggaatattacccttgggatcaagtagagttcogaagaaga 2772
Db 2520 ATTCCGGGAACGAATGCTCGAAATCAACCTTTGGGATCAAGCTCGAGTTCGAGAGAAAGA 2579
QY 2773 gacgcgaattcttagggagagatttaattgaatttggaacacagcttctgtagatgagcc 2832
Db 2580 AAGTGAATTTCTTAGGAGAGATTTTAATTGAATTTGAAACACACTCTGTTGATGATGAAC 2639
QY 2833 gcactgtataagcttgagacccea tga tgcctctcaattgccaactccctccctccccc 2892
Db 2640 ACACCTGTACAAACTTCAGACCCATGATGTCCTTCATTGGCACCTCCCTTCCTTCCTC 2699
QY 2893 atatctgccccggagagacatccatgagagagcccaagcgcaagctgcgaaggtcgaa 2952
Db 2700 ATATATGCCACGGAGACAGCTGCATGAGAGAGAGGCCCAACACGAGGTTGCAAGAGTCAA 2759
QY 2953 gagaataagtgagaaatgagatgctgactacagctgaggaatgagctgaggatgagtc 3012
Db 2760 GAGAATTAAGTGAACGAGCAATATCCGACTGTGACTGGAGAGACGGCTGGAGTGGTTC 2819
QY 3013 agattatcgacacaatgagcgagatcctcaaaagctccagctgtgcgtgccaagaaga 3072
Db 2820 AGATTACCGACAGATGGCGGTGATCTTCAAAAGCTCCACATTATCAGTGCACAGAACAGT 2879
QY 3073 catgtatcaaatcaatctgctcaaccatcaaggtctcctcatcgaatgaatgtlataggaag 3132
Db 2880 CATGTATCAAAACATTTGCTACCATCAGGCTCCTCATCGAATGATGTTATAGGAAG 2939
QY 3133 gacaagtgatggtgcgctatgctccctcctctcaaaagaaatgctggaacagggacag 3192
Db 2940 GACTAGTATGATGCTGCGCTAGTGTCCCTCCTCCTCAAGAAATGTGAAACAGGGGGCTTGC 2999
QY 3193 agggacagctgtcacttggccattcaacacaatlaagccgaatgtagaaccggtgcat 3252
Db 3000 AGGGACACGTCGCTACGCGCATTTACAAATTAATTAAGCGGATGACAGACACCTGTGAT 3059
QY 3253 ggaagccaactactcttcagatagagac----- 3280
Db 3060 GGAGCACCATTACTCTTCAGAGAGACAGTCAATTTCTCACTACTACCTGCTTCGACA 3119
QY 3281 -----aggagtttgaagcagcaagtaaga 3306
Db 3120 CAGGCAGACCATGAGCATACCAACAGGATGGAAGGATTTGTAACAGCAGATGAACA 3179
QY 3307 gccatactcaagatccagatccaagaacaagcgccctcctctagagcggaaccaaccgg 3366
Db 3180 GCCATATCAAGATTCAGATCAACAGAACAGACAGGCGCTCTCTAGAGGGACACACCGCG 3229
QY 3367 ctccagaatcccttgaaagctcctgatacaaaactcaatgaggtcgatgctcctcaatgac 3426
Db 3240 CTCAGATCTCTGAAACGTCGTGATCAAAACCTCATGAGGTGATGCTTCATTAATGAC 3299
QY 3427 tggaaatctgcccctcctcaactccttaataaggtatccacactgtaaccggtctgt 3486
Db 3300 TGGAAATCTGCCCTCTCTCACCTCTTATGAGGTCTCACCCCTGCACTGGCTTGT 3359
QY 3487 ccagaagaacccaatcaagtaactccggaagaagaagggccgaagatccacaagct 3546
Db 3360 CCAGACCACTCGTCAAGTAATCTCCGTAACAGSACCAAGSGGCGGACGTTCTCTACGT 3419
QY 3547 tccaccaagggaaacttggagagaa----- 3572
Db 3420 CCACCAAAAGGGAACACTGGAAGAAATGATCACAGAAAGATGACTCCACAAGAGAGAG 3479
QY 3573 -----gtgctatgatatagaggagaaatcgccaaatgaacttaacaataca 3624
Db 3480 AAATCTAGGTGCTATGAGATTAAGAGAAATTCGCCCAATGAATACTTAACAAATTAACA 3559
QY 3625 aacagtgacgagatccagaccacagactgagcaagatttaccattcgaagatcagctcag 3684
Db 3540 ACAGTATGCGGATCGATCGATCCAGACTGGAACAAAGATTACATTCCAAGTATGCTCAGG 3599
QY 3685 atggatccacaatgaagggcagatactgtttccacitlaaattcctcgyacagtgatgtaa 3744
Db 3600 ATGCAATCCACTAGAGGGGAGATGACTGTTCTTACTTAATCTCTGACAGTGAATGTAAG 3659

| | | | |
|----|------|--|------|
| OY | 3745 | tgatgatctgcggtttccaagactagtagtgcttcttcgtttcaagcgcacagctacat | 3804 |
| Db | 3660 | tgatgtatcttgcgggtttccaaagactagtagtgcttcttcgtttcaagcgcacacgtctacat | 3719 |
| OY | 3805 | gtccgtccaatcaagacgagccagagagaaacagaaatcaagtgctttcatccaatat | 3864 |
| Db | 3720 | gtccgtccaatcaagacgagccagagagaaacagaaatcaagtgctttcatctccaatat | 3779 |
| OY | 3865 | gcaaaacagacagatctggcgctgttcgggagaagaactctgaccaaaagccacagactcagctg | 3924 |
| Db | 3780 | gcaaaacagacagatctggcgctgttcgggagaagaagaactctgaccaaaagccacagactcagctg | 3839 |
| OY | 3925 | agacatctgtctactctgagagaagaatgacgagcaagctccgacactgacgtggcgccct | 3984 |
| Db | 3840 | agacatctgtctactctgagagaagaagaatgacgagcaagctccgacactgacgtggcgccct | 3899 |
| OY | 3985 | gggtaccagctggccaagaagccgagctacagcatctgggcccacaaatgtaactatgtctgtg | 4044 |
| Db | 3900 | gggacacacagtgccaaagaaacggcgagactctgacattggggcccaaaatggtactattgttgg | 3959 |
| OY | 4045 | tcctccacgagaaagctgcagctgctcctcaactcagccaaacccgaagagaggttaaaaa | 4104 |
| Db | 3960 | tcctccacgagaaagctgcagctgctcctcaactcagccaaacccgaagagaggttaaaaa | 4019 |
| OY | 4105 | gctacgagacactgtctcagaagaacagagacgggctagcagctggagaatgaagaaactg | 4164 |
| Db | 4020 | gctacgagacacactgtctcagaagaagaacagagacgggctagcagctggagaatgaagaaactg | 4079 |
| OY | 4165 | gattagcccgccagccagccagcggaatcccaacagctggcagatgaacgctatagctcaga | 4224 |
| Db | 4080 | gattagccacacagccagccagcggaatcccaacagctggcagatgaacgctatagctcaga | 4139 |
| OY | 4225 | aggaatctcgaactctcccccggggctccgcttgagctctgacagccagctcgaatgttccct | 4284 |
| Db | 4140 | aggaatctcgaactcttcttcctgggggtccgctctgacagccagctcgaatgttccct | 4199 |
| OY | 4285 | ggaatgcctggccctgctcactcagctagctgggacgcacagccctggtctactcctcgaatgg | 4344 |
| Db | 4200 | tgatggtctggcccttcgctcactcagctagctgggacgcacagccctggtctactcctcgaatgg | 4259 |
| OY | 4345 | tgaccttaaggtggggaatgatatgataaaaaagagacagcttggaagtgaatatccatccgggc | 4404 |
| Db | 4260 | tgacatctcaggtggggatgatgatgataaagagagacagcttggaagtgaatatccatccgggc | 4319 |
| OY | 4405 | ggcgagcccttggtgtaaaaaacaggtctccaagacactgcgcagccagctgatactcaagtgtga | 4464 |
| Db | 4320 | tcgagagcccttggtgtaaaacacaggtctccaagacactgcgcagccagctgatactcaaggtgta | 4379 |
| OY | 4465 | tcgtgtagaacacggaagctctgataagccaaaaaagaaacccaagctggcgagaagaacctt | 4524 |
| Db | 4380 | cctgtttagacacacgagctgtgcatacgcacaaaagaaacccaagctggcgagaagaacctt | 4439 |
| OY | 4525 | ggagcccccgtatccagcaagctctctgtcccttcgagggagagcccccagggggaggtgttaca | 4584 |
| Db | 4440 | ggagcccccgtatccagcaagctctgttaccttttcgaaagagagcccccaggggaaaggtttaca | 4499 |
| OY | 4585 | gatactgtctgggagagactatagctgtatgtgatacacaatccctttatggagatggcgca | 4644 |
| Db | 4500 | gatactgtctgggagagattatagctgtgtatgtgatacacaatccctttatggagatggcgcca | 4559 |
| OY | 4645 | gatactcttagatgaactctgaaactatccacaatcagtgatgtatgtatgtgttcaaaactctccc | 4704 |
| Db | 4560 | gatactcttagatgaactctgaaactatccacaatcagtgatgtatgtatgtgttcaaaactctccc | 4619 |
| OY | 4705 | tcctctccctccctagtagataccaaactcgggacactctgacaaagaagagcttcccaatcgctc | 4764 |
| Db | 4620 | tcctctccctccctagtagataccaaactcgggacactctgacaaagaagagcttcccaatcgctc | 4679 |
| OY | 4765 | tcctggaagcttaccagagaccttcttaccctgtttcataagcaactataaaactgttca | 4824 |
| Db | 4680 | tcctggaagcttctactggacacttcttactctcgtttatagcaactatataaaactgttgcata | 4739 |
| OY | 4825 | caacacacgagctatacaaaaacagagaagaaacgcacagcttggaagccctctgttaacact | 4884 |

| DB | 4740 | CAACACACGCGTTTCAAAAACCAAGAAACACACCACAGGTGG-6GACCTGTGTCCTCACT | 4798 |
|------------|---|---|-------------|
| OY | 4885 | gcatcttgatgtgtgtgtctatagaagccacagctcattagatacaagaagtcctgtgtc | 4944 |
| DB | 4799 | GCATCTTGATGTGTGTGTCTCTCCGAGGCCACAGCTCTAGGGATTAAGAGCCATCTGTCTC | 4858 |
| OY | 4945 | tcagagaaga---tcgtacacatgtgtccctagcaaaagg | 4980 |
| DB | 4859 | TCAGAGGAAGCCATCCGTGCACAGTGTGCCCTAGCAAAAGG | 4898 |
| RESULT | 5 | | |
| LOCUS | AF199324 | 5172 bp | 04-JUL-2000 |
| DEFINITION | Rattus norvegicus RIM2-2B (Rim2-2B) mRNA, complete cds. | | |
| ACCESSION | AF199324 | | |
| VERSION | AF199324.1 | GI:8925861 | |
| KEYWORDS | | | |
| SOURCE | Norway rat. | | |
| ORGANISM | Rattus norvegicus | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. | | |
| AUTHORS | 1 (bases 1 to 5172) | | |
| TITLE | Wang, Y., Sugita, S. and Sudhof, T. C. | | |
| JOURNAL | The RIM/NIM family of neuronal C2 domain proteins. INTERACTIONS WITH RAB1 AND A NEW CLASS OF SRC HOMOLOGY 3 DOMAIN PROTEINS | | |
| MEDLINE | J. Biol. Chem. 275 (26), 20033-20044 (2000) | | |
| AUTHORS | 2 (bases 1 to 5172) | | |
| TITLE | Wang, Y. and Sudhof, T. C. | | |
| JOURNAL | Submitted (27-OCT-1999) Center for Basic Neuroscience, The University of Texas Southwestern Medical Center, 6000 Harry Hines Blvd., Dallas, TX 75235-9111, USA | | |
| FEATURES | location/Qualifiers | | |
| source | 1..5172 | | |
| gene | /organism="Rattus norvegicus" | | |
| cds | /db_xref="taxon:10116" | | |
| | 1..5172 | | |
| | /gene="Rim2-2B" | | |
| | 52..4251 | | |
| | /gene="Rim2-2B" | | |
| | /codon_start=1 | | |
| | /product="Rim2-2B" | | |
| | /protein_id="AAF81646.1" | | |
| | /db_xref="GI:8925862" | | |
| | /translation="MSAPLGGRRPAPTPAASOPPEMPDLSHLTEERKLIQAVM DROKKEBEKEQSVLKLHQEEMKEDVKKKKGESQOEQEGKAPFGICJHKTFAD GKNGCSCYOTKPCARGGRVSLRSNNVMVNCICRQOELTSLGAMFVNSGNTPO QPDQKALRGLAESAPQEKRAKLEONFOFGPPGDSVPAVERAAGLTRDSTIKG SGMKHQIASMPSPKRSFVSVRQNNRYDQSEREYSQVPSDSTMPSPSYADR RSQEPQYEEDPLNTRDSNRKGRHSKKEIVDEDEVSEDEYEROREEYQVARY SDMPARYPVKQPEYEQMRHAIVSASRAHRRSDVSLANLELDSRI SLRLMDPS RQBSVERRAMENORSYMERTEAGOGSSYPORTNHPPPRRAPRI PLDRBELRR ADSLRKHQHLDPSSAVRTRKREKMETMLRNSLSDOSESVRRPVPKSKGKKMR OVSLSSREELASPEYTCSDVTEISESCEKEDSKGKRKTSQEQAVMDSNTRSR OKKMYGGGSHLDELENSERQTRDSCVDTCSTTLWERSHSEHYVMQSPGDGR LIGLILNLRKIDGSVPDSCAMGLVYGGKMTSEGRALFTYKVGSLADLVGRL RPGDEVLEWNGRLQOGATFEVYNIILSKKEPQELVSPHIGDMPIRDPSTHAQLE SSSSFSQKMDRPSISVTSFSPGMLRDVQFLSGQLSLKMRDQGHOLIYILGA KDLCSFREDGRPNRYVYKTYFLPDRSDKNKRTKTKTLEPKNMOTFISVPHRRER ERMLEITLMQDARVRESESPICETILLETALLDDEPHMYKLOTHDVSSILPHPS YMPRQLHGESPTLRLORSKTSIDSEVSDIDCEGVGVSLGRHGDLDLSTLSVEE OVMSNHCSPSPGPHRVDIVIGRTSWSBSPVPPQNRNVEQLGTRAGHNTLSRMDR HRVMDHYSESERDSHFLPLRPSRRROTSEHHHRGDRCEADARQPYHRSSTEQRPILL ERTTTRRSERSRADNTLIRMSWPSLMTGRSAPSPALSRMPTGTSVOTSSPVTGR SGEONLFPKGTGLERGGGKKLRSTYGRSTETGALVSRMMMTPOASRESTDSMNY SSEBNLFPKRYRLASDSQSFSDLDGLGPAOLVGNQTLATAMEQIYQGMDDKQOLEY EITLRAGLVVAEGSKTLPAPYKAVYLLDNGCLAKKTKTAKRTLELYTOOLLSFEES PQGVGLIIVGDDGRMDHKSFMGVAAIILDELELSMWVIGWPKLPPSSILVPTPLAP LTPRAASQSLSESTGSPSYRS" | | |

BASE COUNT 1555 a 1190 c 1286 g 1141 t
ORIGIN

Query Match 64.2% Score 3197.6; DB 94; Length 5172;
Best Local Similarity 81.2%; Pred. No. 0;
Matches 4101; Conservative 0; Mismatches 239; Indels 710; Gaps 5;

OY 1 gcttcctaaggtgtgttcgcgtccgcaccaacatgltcgcgtccgcgtcgggcccgggcgcg 60
|||||
Db 21 GCTTCCTAGGGGTGTGTGGTCCACCACAAACATGTGGCTCAGTCCAGTCCGCCCGGGCGCG 80
|||||
OY 61 ccgggtcccaaccccgcgcgccctccaacctctccgcgaccccgagatgcccgaactcag 120
|||||
Db 81 CCGGCTCCACCACCCCGGGGCGCTTCACAGCTCTCCCGCGCCGAGATGCGGACCTCAG 140
|||||
OY 121 ccactcaggaagaagagagaagaatactcgtgctgltcagatcgltcagagaagaaga 180
|||||
Db 141 CCACCTCAGGAGAGAGAGAGAGAAATCATCCAGGCTGTCTATGATCGTCCAGAGAAAGA 200
|||||
OY 181 agagagagaagagcagtcgcgtgltcaagatcaagaagaacaacaagcaaacccagaca 240
|||||
Db 201 AGAGAGAGAGAGAGAGTCCGTCTCAA----- 227
|||||
OY 241 gtggttcccttagtgggatactcgaactggttaaatgaatctgcagccccagcaaaa 300
|||||
Db 228 ----- 227
|||||
OY 301 acaacccaatgagaagagccccaagaacatgltcaccacaatttgaatatgaaga 360
|||||
Db 228 -----AAAGCTGCATCAACAMTGAATATATAAGA 260
|||||
OY 361 gcaagtcagaagaatgagagagaatcgcagcaagcaagaagcagaaggtgatagcccc 420
|||||
Db 261 GCAAGTCAAGAAATGGGAGAGAAATCACACAGCAGCAAGACAGAGGCGCAGCGCCC 320
|||||
OY 421 gacctgtgcatctgcacaagaacaatttgcagatgagatgagcgcacataattgltcta 480
|||||
Db 321 GACCTGCGGCATCTGCCACAGACAAAATTTGCAGATGATGGGCGCATACGTTCATATA 380
|||||
OY 481 ttgccaaccgaagtctgtctcgatgltggaagltcgagltcttlaagctcaaacgaagt 540
|||||
Db 381 TTGCCAAACCAAGTCTGTGCTGTGTGGAGGTGAGTGCATATACGCTCAAAACAAGGT 440
|||||
OY 541 tatgtggtgtgtaatttgtgcgcaaaaacaagaanaattcctaataatcagagatg 600
|||||
Db 441 TATGTGGGTGTAAATTTGTGCCGAAACAAACAAGAAATCTCAATAAGTGGCGGCGTG 500
|||||
OY 601 gtttataatagtggtctaaacacactgcagcaacctgatacaaaagtctcctcgaggct 660
|||||
Db 501 GTTTTACATATAGTGGTCTATATACCCGCGACGAGCTGATCAAAAGGCTCTGTGAGGGCT 560
|||||
OY 661 tcgaaatgagaagccctcctcagaagaagaacaaacatacagagcagccccagtcca 720
|||||
Db 561 TCGCAGTGAAGAAAGCCCTCAGAGAGAAAGAGCAAAATCATGATGACAGACGAGTTCCA 620
|||||
OY 721 agagagcccaagtgacttatcaagtaactgcaagtgtgagaagaagccgaagctcaaggtcac 780
|||||
Db 621 GGGACCCCGGGGACTCATCACTACTGAGTTGAGAGAGAGTGAAGTCTATGGGCTCAC 680
|||||
OY 781 aagacagagatactataaataatgatacgaagtgaaagcacacagatltccagtgacatgcc 840
|||||
Db 681 AAGACAGGATTTCTATTAAGAAATGGCTCAGAAATGAACACACAGATCGCACTGACATGCC 740
|||||
OY 841 ttcaagacagaaaacgaagltccatcaagtgtccagggatcaaaatctgaagatcagcgcaag 900
|||||
Db 741 TTCAGACACAAAAAAGAAAGTCCGTAGTCCAGGATCAAGGATCAAAATGGAATVAGCAACAAAG 800
|||||
OY 901 tgaagaagaagagagactcacagatagtcttcctcagatgtaacaatgcaagaatctcc 960
|||||
Db 801 TGAAGAAAGAGAGGAATATTCACAGTATGTTCCTTCAGATGACAAATGCTAGATCTCC 860
|||||
OY 961 ttcgagatagtcatagacgactcagcgtgagccctcaatttatgaagaacctgtgtca 1020
|||||

|||||
Db 861 ATCAGATTATGCTATATGAGCATCTCAACGTAACCTCAATTTTATGAAGAACTGATCA 920
|||||
OY 1021 tttaaatltaacagggatcttaacagagaagccatagacatctccaaagatattgttga 1080
|||||
Db 921 TTTAAATTAATAGGATTTCTTAACAGAGAGGGCCATAGACATTTCCAAAGATATATTTGTAGA 980
|||||
OY 1081 tgaagaagaatgtagagcagaagatgaaatgaagaagacaagaagagaagaaatacca 1140
|||||
Db 981 CGAGGAAGATGTGAGACACAGATGAATATGAAAGACAAAGACAGAGAGAGATGACCA 1040
|||||
OY 1141 ggcacgtacagaagltgataccaatctggcccggtatcccgtaaaagccaacctcaga 1200
|||||
Db 1041 GGCAGGTACAGAAAGTGTATCCAAATTTGGCCGGTATCCGGTAAAGCCACACACCTATGA 1100
|||||
OY 1201 agaaacaatgcgcatccagcttgaggtgtlccaggtcagagacatgagagaagcagaatga 1260
|||||
Db 1101 AGACCAATGCGGATCCACGCTGAAGTGTCCGGGACACACGAGAGAGACACAGTGA 1160
|||||
OY 1261 tgttctcttgcaaacgctggaactagaagatccagatcttcctgctaaagatgagatag 1320
|||||
Db 1161 TGTTCCTTTGGCAAAATGCTGAACCTGAAGATTCAGAGATTTCTGTGCTAAGATGATAG 1220
|||||
OY 1321 accataaggaagaatcgtatctgaacgtagagctgcaatggaataaccaagatcgtga 1380
|||||
Db 1221 ACCATCAAGGCAAAAGATCTGTATCTGAACGTAGAGCTGCAAATGGAAMCCAGGATCTGA 1280
|||||
OY 1381 ttcaatggaagaagactggaagagctcaaggagaagaaglttltccacaaggacctcca 1440
|||||
Db 1281 TTTCAATGGAAGAAACTCAGAGAGCTCAGGACAAAGTTTTCATCCAAAGAGACCAAMA 1340
|||||
OY 1441 tcaatgctctcccaaccctcgcgcgagagccctaataccgttgtatagaaccaatgaagcg 1500
|||||
Db 1341 TCATGTCTCTCTCTACCCCTCCGAGAGGCCCTATACCGCTCGATAGACACAGAGCTGAGCG 1400
|||||
OY 1501 cgctgaatccctcaggaagaacagcacaacttaagatcccaagctcgtcgttgaagaagaacgaa 1560
|||||
Db 1401 TGCCGACTCCCTACCGAACAACACACCACTTAATCCACTCTGCTGTGAAGGAAGAAAGAA 1460
|||||
OY 1561 gcgagaanaaatggaacacatgttlaaggaatgttcttgaagltcagaaccagtcagagtc 1620
|||||
Db 1461 GCGAGAAAAAATGGAACAAATGTTGAGGAATGATTTGTTGAGTTGACAGCACAGCTGAGTGC 1520
|||||
OY 1621 agtgaagcgcgccccaacgaagcctcctaataatccagaagaagagttaaaatgcgcgaagt 1680
|||||
Db 1521 AGTGAAGCGCGCCCAACCAAGGCTCATTAATCCAGAAAGAGGTAAATGCGCCAGGT 1580
|||||
OY 1681 ttcaactgagcagctcggaggaagagctgtgcatccacacctgaatatacaagctgtgata 1740
|||||
Db 1581 TTTCACTGAGCAGCTCAGAGAGAGAGTTGGCATCCAGGCTGAGATATACAGCTGTGATGA 1640
|||||
OY 1741 tgtgagctggaagcgaagagtgtagtgaagaagggagcagtcacaaagggaaagaa 1800
|||||
Db 1641 CGTGGAGATTTGAAGAGAGAGAGCTAGTGAAGAAAGAGACAGTCAAAAGGAAAGAA 1700
|||||
OY 1801 aactagtgagcagggagtttgcggaatcttaacccaagctcagagagaacaaagaag 1860
|||||
Db 1701 AACTAGTGAAGCAGCGAGTTTGGTGGACTCTAACACACAGCTCTCAGAGACAAAAAGAAAGT 1760
|||||
OY 1861 gatgtactatggtggccactcttltgaaagaagatltggaaatgtctcgaagctcagataa 1920
|||||
Db 1761 GATGTACTCTGTGGGCGCACTCTTTGACAGAGATTTGGAAATGCTGACGCTCAGACTAA 1820
|||||
OY 1921 ggaactctgggtlagatacctgtagtgaacaaaccttaacgagaagacagataagtg 1980
|||||
Db 1821 GGACTGTGGGTAGACACCGTGTAGACACACCTTAACGAGAGAGCGTGAACCTATGTGA 1880
|||||
OY 1981 taagcaacctgtgacctgagcagccatccaaagaatgagatcgctcaatttggctatctt 2040
|||||
Db 1881 TGAAGCACCTGTGACCTGGCAGCATCCAAAGATGAGATGCGCTTAATGTGCTATATTT 1940
|||||
OY 2041 attaaataagcgtttaaagaatggaggtgtacctcgaagatcaaggaacaaatgctgggctt 2100
|||||

Db 1941 ATTAAATGAAGCGTTTAAAGATGGAGTGTACCTCGAATTCAGAGCAATGCTGGCCTT 2000
QY 2101 aaaggtgttagagaggaagaatgactgaatcaggtcagcttctgtcatttattaccaaagt 2160
Db 2001 AAAAGTTGAGGAGGAGAAAGTGAAGTCAATCAGGTCGACTTGTGCATTTATTTACCAAAAGT 2060
QY 2161 aaaaaaagaaggtttagctgatactagctgaagacatctagaaccaggtgatgaagcttga 2220
Db 2061 GAAAAAGAGAGTGTAGCTGATCTGTAGSACATCTTAGACACAGTGATGAAGCTTGTGA 2120
QY 2221 atggaaatggagagctatctgaagaagccacatttgagaagattacaacatctataga 2280
Db 2121 ATGGAATGGAAGGCTATTGTGCAAGAGAGCCATTTGAGAAAGTTTACAACTTATTTCTGGA 2180
QY 2281 atccaaacctgaaccacaagtltgaagctgtgttttcaaggccaattggaatttccctag 2340
Db 2181 ATCCAAAGCCCTGAACCAACAACTTGAAGCTTGTGTTCAAGGCCGATTTGAGATGAGTCCCTAG 2240
QY 2341 aataacctgaatgaacgcatgcacaaactggaatccagttcagcttcaagcttgaattcca 2400
Db 2241 AATACCTGATAGACATCATATCAGACAACTGGAATCCAGTTCTAGCTCATTTGATCTCAAAA 2300
QY 2401 aatgagcgcctcctctatatacgcgttaccccaacaaatgaagtcggaatgcctgaaggaatgt 2460
Db 2301 AATGGAATCGCTCCCTATATATCGGTACTCCCTCCATGAGTCTGGAATGCTGAGAGGATGT 2360
QY 2461 ccgcagcttctatctggaagacttcaataaactatgtttgacaaggttggtaacca 2520
Db 2361 CCACAGCTTCTATCTGAGACAGCTTTCATTAATAACTATGTTTACAAAGTTGCTGACCA 2420
QY 2521 gttagtagtatacaattttggagacaaagatctccctccaggaagaatggagagccag 2580
Db 2421 GTTAAATAGTTACATTTTGGAGACAAAGATCTCCCTTCCAGGAGAAATGGAGGCCAAG 2480
QY 2581 gaatccttatgttaagaattacttccctccagabagaagtgtataaaataagagaagaac 2640
Db 2481 GAATCCTTATGTATAATTTACTTCCCTCCAGACGAAGATGATAAAAACAAGAGAAGAAC 2540
QY 2641 aaaaaaagtcagaaaaaactttgaaacccaatgaaaccaagacttcaatttacttccgt 2700
Db 2541 AAAAAAGTCGAAGAAAACCTTTGGAACCCAAATGGAACCAAGCTTTCATTTATCTCCCTGT 2600
QY 2701 ccacccgaagaagaatccgttgaacgaatgctggaatcaaccccttgggaatcaagctagag 2760
Db 2601 TCCACCAAGAGATTCOGGAGACGAATGCTCGAAATCAACCCCTTGGGATCAAGCTCGAGT 2660
QY 2761 tcgagaagaagaagcgaatctcttagagagaatttaattgaatttgaacagcttgc 2820
Db 2661 TCCAGAGAGAAAGAAAGTGAATTCCTAGAGAGATTTTAAATTGAATTGMAACAGCTCTGTT 2720
QY 2821 aggtatgtgacgcgactggtatagctgcagagaccatgaatgtctcccatattgcaactccc 2880
Db 2721 AGATGATGAACCACTGCTGTAACAACTTCAGACCAATGATGTCCTCATTTGCCACTCC 2780
QY 2881 tgcaccttcccatactgcgccggagcagctccacggagagagcccaacgcgcagct 2940
Db 2781 TCCACCTTCCCATATATATGCGACGAGACAGCTGCAATGGAAGAGCCCAACACGAGGTT 2840
QY 2941 gcaaaagtcgaagaagaataagtgacagtgaaagtgtctgaactagacgtgcagagatgagcgt 3000
Db 2841 GCAAAGGCTCGAAGAGATTAAGTGAACAGCAAGTATCCGACATATGACTGCGAGAGCGCGT 2900
QY 3001 gggagagatgtgagattatagacacaaatggccgagatcttcaaaagctcaagttgtgctg 3060
Db 2901 GGAGAGCTGCTCAGATTAACGACACGATGGCCGATGCTTCAAAAGCTCCACATTAATACGT 2960
QY 3061 gccaagaacaagtcatgtcatcaaatcatgtctcaaccaatcaaggtctccctcatcgagtaga 3120
Db 2961 GCCAGAACAAAGTATGTCATCAAAACCATTTGCTACACATCAAGGCTCTCCTCATTCAGGATGA 3020
QY 3121 tgttatagaagaagaaggtcatgtatgtctgctagtgtgcccttccctccaaagaagaatgtga 3180
Db 3021 TGTATTAGGAAGGACTAGGTCATGATGCTGCTAGTGTCCCTCCTCCTCAAAAGGAATGTGGA 3080

QY 3181 acaggggacccgaaggagacagctgtctactggccattacaacaaatltagccgaatgtatag 3240
Db 3081 ACAGGGGCTTTCGAGGGACACAGTGTCTACTGTGCGCATTTACAAATACATTAAGCCGATGGACAG 3140
QY 3241 acacgcgtcatatgaatgacacactactcttcaagtagagac-----aggatgtgaagc 3280
Db 3141 ACACCGTGTCATGAGACGACCATTAATCTTTCAGAGAGAGACAGATCAATTTTCTCACTATAC 3200
QY 3281 -----aggatgtgaagc 3294
Db 3201 TCGCTTCACACAGCAGACAGACAGTGCATCACACACAGCATGGAAGGATTTGTGAAGC 3260
QY 3295 agcagaatagaacacacatatcacagatccagaatacaagaacaaagcctctcccaagcg 3354
Db 3261 AGCAGATAGACACCATATTCAGATTCAGATTCAGATTCAGACGAACAGGCCCTCTCTAGAGCG 3320
QY 3355 gacacacaccgcctccagatctccttgaagcttccgtatatacaaacctcatagatgtatgac 3414
Db 3321 GACACACACCCGCTCCAGATCCTGTGAAGCGTGATACAAACCTCATGAGGTGATGCC 3380
QY 3415 ttaataatgaactggaagaatctgcccctctcaacctgtacctatcgaggttcaacctcg 3474
Db 3381 TTTCAATATGACTGGAAGATCTGCCCTCCTTCACTGCTTATCGAGGTCTCACCCCTCG 3440
QY 3475 taaccggtctgtccagacaaagcccatcaagttactccgaggacaaagagggccgaca 3534
Db 3441 CACTGCTCTGTCCAGACCAAGTCCGTCAATTAATCTCCGTACAGAGAGAGGCGGCA 3500
QY 3535 gcttccacagcttccacaaaggaaacttggagagaagtgtctatgtatatagagagag 3594
Db 3501 GCTTCTCAGCTCCACCAAGGAACACTGGAAG----- 3536
QY 3595 aaatgcacaaatgaaacttaacaatacaaaacaggttagccgagatcagaacccagactgga 3654
Db 3537 ----- 3536
QY 3655 gaaagattaccatctggaagtatgcgtcaggaatggatccacatagagggcagatactgt 3714
Db 3537 ----- 3536
QY 3715 ttccactaaatccctcgcgacagtgatgaatgtatctcgcggtttcaagactagtag 3774
Db 3537 ----- 3536
QY 3775 tgcctctgtttcagacgaacaaagctacatgltccgtccaatcagaagcgccgagaagaaa 3834
Db 3537 ----- 3536
QY 3835 caggaaatcatgltcttcatccaataatgcaaaacagacagatlggcygtcggggaa 3894
Db 3537 ----- 3536
QY 3895 gaacttgaccaaaagacacagcatcagtgagagacatgtgtcacttggagaagaaatgacgy 3954
Db 3537 ----- 3536
QY 3955 cagcagctcagacatgcagtgaggcgccctgggtacacagtggcaagaagcgcgatctag 4014
Db 3537 ----- 3536
QY 4015 catltygggccaaaatgtgtagctatgtgtgtctcctcaagaaagctgcagtgctctca 4074
Db 3537 ----- 3536
QY 4075 actcagccaaaacgaaggaggaatgaaagctacgagacatgttctagagaagacagga 4134
Db 3537 -----AGAGAGAGGTAAAGGCTACGACGACGCTCTCCAGAGAACGACAGGA 3581
QY 4135 gaccgggtcagcagtggaatgtgaagaaacttgatgacccgccaagccagcggaatccac 4194
Db 3582 GACCGGGCTAAGCAGATGAGATGAGAACTGATGACAAAGACAGCCGCGGGAATCCAC 3641

Db 228 -----AAAGCTGCATCAACAAATTTGAATGTAAGA 260
OY 361 gcaagtcaagaagatgaggaggaatcgcagcagcagcaagagcagagatgagtcgcc 420
Db 261 GCAGATCAAGAAATGGGAGAGAAATCACAAACAGAGACAGAGCGCGCGCC 320
OY 421 gacccgtgcacatcccaagaagaatttgcagatgagtcggccataattgctcta 480
Db 321 GACCTGCGGCATCTGCCACAGACAAAATTTGCAGATGATGCGGCCATTAATCTTATA 380
OY 481 ttgcacaaccaaattctgtcgtcgcagatgtagaggttcgagtgcttcaacgtcaaaagt 540
Db 381 TTGCCAAACCAAGTTCTGCTGCTGTTGGAGGTGAGTGTCAATTACGCTCAACCAAGGT 440
OY 541 tatgtgtgtgttaatttctgcgcgaaacaagaatactcaactaataatcagagcatg 600
Db 441 TATGTGGGTGTAAATTTGTGCGAAACCAACAAAGATCTCACTAAGTGGCGGTG 500
OY 601 gtttataatagtggtctcaacacactgcagcaacctgatacaaggctcctcgaggct 660
Db 501 GTTTTACAAATAGTGGGTCTAATACACCGCAGCGCTGATCAAAAGGCTCTTCAGAGGCT 560
OY 661 tcgaaatlgagaagccctcagagagaagaagaacaaactacagagcagcccgatcca 720
Db 561 TCGCAGTGAAGAGCCCTCAGGAGAAAGCAAACTCATGAGCAGACGAGTTCCA 620
OY 721 aggaagcccaagtgactatcagctacgtcagtgtagaagagcccgagatcagtggtcac 780
Db 621 GGGACCCCCGGGTGACTCTCACTACCTGCGAGTGGAGAGGTGAGACCTCATGGGCTCAC 680
OY 781 aagacagagatcatttaaaatlgatcagagagtgaaagcaccagatctccagtgacagcc 840
Db 681 AAGACAGAGTTCTATTAAAGATGCTCAGGAATGAAACACCAATGCCAGTGCATGCC 740
OY 841 ttcaagacagaanaacgaagtcacatcagtgctcagagatcaaaatcgaagatcagacaag 900
Db 741 TTCAGACAAAAGAAAGAGTCCGCTGAGTGTCCAGGATCAAAATGGAATGAGACAAAG 800
OY 901 tgaagaagaagagagactactcaagatagtctcctcagatgtagtaaatgccaagatccc 960
Db 801 TCAGAGAAAGAGGAATATTACAGATATGTTCTTCAGATAGCAAAATGGCTTAATATCCC 860
OY 961 ttcgagatctgcgatagacgatctcagcgttagagcccaatttataagaagactgtcca 1020
Db 861 ATCAGATTTTGCATAGACGATCTCAAGCTGAACCTTAATTTATGAAGAACCTGATCA 920
OY 1021 tttaaaatcaaggatcttcaacagagagagccatagacatcccaagaaglatatgtgga 1080
Db 921 TTTAAATTTAGGATTTCTAACAGAGAGGCCATAGCATTTCCAAAGATATTATTGTAAGA 980
OY 1081 tgaatgaagatgtgagagcagagatgaataatgaagaagaagaagaagaataacca 1140
Db 981 CCAACGAAGATGTGGAAGACGAGATGATATGAAGCAAAAGAGAGAGGAGGATACA 1040
OY 1141 ggcacgctcacaagaatgataccaaatctgcccgttatcccgtaagaagcacaaccttaca 1200
Db 1041 GGCACGCTTCAGAGATGATCCAAATTTGGCCCGTATCCGTTAAGCCACAACTTATAGA 1100
OY 1201 agaaacaatgycatccacagcttagaggtgtccagagcagacatagagaagaagcagatga 1260
Db 1101 AAGAGCAAAATGGGATCCAGCGTGAAGTGTCCCGGCGACGACAGAGAGAAGGACAGTGA 1160
OY 1261 tgttctcttgcaaacgctggaactagaagaatccagaattctctgtaagaatgagatag 1320
Db 1161 TGTTCCTTTGGCAAAATGCTGAACCTAGAAAGATTCACAGATTTCTCTGTAAGAGATGAGAG 1220
OY 1321 accatcaagaagcaagaatctgtatctgaagctagagctgcaatgtaaaaaaccaagatgta 1380
Db 1221 ACCATCAAGGCAAAAGATCTGTATCTGACGTAGAGCTGCAATGGAAGAACGACGCACTCTTA 1280
OY 1381 ttcaatlgaaagaagactcagagaggtcctcagagacaagttctlatccaaaagagactcaaa 1440
|||||

Db 1281 TTCAATGGAANAAGACTCGAAGAGCTCAGGACAAAGTCTTATCTACCAAAAGGACACACAA 1340
OY 1441 tcatagctccacacccctcgagagccctataccgcttgatagaccacaatgagcg 1500
Db 1341 TCATAGCTCTCTACCCCTCGCAGAGGCCCTATACGCTGTGATAGACCAAGCTGAGAGCG 1400
OY 1501 cgtgacccctcagaaacagcaccatlagatcccaagctctgctgtagaagaagaa 1560
Db 1401 TGCCGACTCCCTACGGAACCAACACACACACTAGATCCAGCTGTGCTTAAGGAAGAAAGAA 1460
OY 1561 gcgagaaaaaatgynaacocatgltlaaggatlgatctcttgatltcagaaccaagtcagtc 1620
Db 1461 GCGAGAAAAAATGAAACCAATGTTGAGGAATGATTTCTTACGTTCAGACCAAGTCTGAGTCT 1520
OY 1621 agtggagcgcccccacaagagctcataatccaaagaagaggttaaaatgagccaggt 1680
Db 1521 AGTGAAGCGCCCCCAAGGCTCATTAATCAAGAAAGAGGATAAATATGCCCGAGGT 1580
OY 1681 ttcactgagcaagctcgagaagagctgagatccacacccgtagtatacaagctgtagta 1740
Db 1581 TTCACGTGACGACGCTCAGAGAGAGAGTGGCATTCACGCCCTGAGATATCAACGCTGTATGA 1640
OY 1741 tgtgagctggaagcagagatgtgagtgagaagaaggagacagtcocaaaaagynaagaaa 1800
Db 1641 CGTGAGATTGAAAGCGAGAGCGTAGGTGAGAAAGGAGACA----- 1681
OY 1801 aactagtgagcgagaggttctgcgatatcaacacaggtctgagagacaaaaagaag 1860
Db 1682 ----- 1681
OY 1861 gatgtactatgtygacacclctltggaagaagattgaaatgltcagctccagatlaa 1920
Db 1682 -----TGAGTACAGCTGTTGGAGCAGCATGGCTCTTG 1712
OY 1921 ggaacttggygtagatacctgtgtagcacaaccttaacgaagagcatagccaagtagta 1980
Db 1713 G-----CATAGCAGTGAAGCGCANCCCCAAATGTC 1739
OY 1981 taagaacccgtgacctgagcagccatccaaagaatgtagatcgccctaatttgtagtattt 2040
Db 1740 TTTGACCCCTGTGACCTGGAGCCATCCAAAGATGGAGATGCCCTAATTTGGCTATATTT 1799
OY 2041 attaaataagcgttlaaagaatggaatgtagtctcagacatcaagagcaatgycggagct 2100
Db 1800 ATTAATAATAGCGTTTAAAGATGGAGCTGATCCTCGAGATTCAGAGACAAATCGCGCCTT 1859
OY 2101 aaaggtgttagagaagaagaatgactgaatcagatcgagcttgtgcatltatlaacaaagt 2160
Db 1860 AAAGGTTGTAGGAGAAAGATGACTGAATCAGGTGCGACTTTGTGCATTTTATACCAAGT 1919
OY 2161 aaaaaaagaagtttagcttgatctgtagacatcttaagcccggtgtagaagcttgga 2220
Db 1920 GAAAAAAGAGAGTTTATGCTGATCTGTAGGACATCTTAAAGCCAGGTATGAAAGCTTGA 1979
OY 2221 atggaatggagagctatctcaagagagccacatctgagaagattlaacacatlatctaga 2280
Db 1980 ATGGAATGGAAGGCTATTGCAAGAGGCAACATTTGAGAGAGTTTACAACATATTCTAGA 2039
OY 2281 atccaaacctgaaccacaagltgagcttgtlttcaagggccaatlgagagatlatccag 2340
Db 2040 ATCAAGACCTGAACCAACAGTTGAGCTTGTGTTCAAGGCCGATGAGATATGAGCTAG 2099
OY 2341 aatccttgtagagcagcagacatgcaactggaatccagttctagctcatltgaaatccaaa 2400
Db 2100 AATACCTGTATAGACATCATGCAACACTGGAATCCAGTTCTAGCATTTGATTCATAAA 2159
OY 2401 aatgacgctccctctatccgttlaacctaccacatgaagctccgtgagtcgagagaggt 2460
Db 2160 AATGATCTGCTCCTTATATACGTTACCTCTCCATAGTCTGTGCAATGCTGAGAGGATGT 2219
OY 2461 cccgcagttcttactgagacagcttccaataaaactatglttgacaaggttgtgtacca 2520
Db 2220 CCACAGTTCTTATCTGAGACGCTTCAATAAACATATGTTGACAAAGTTGGTACACA 2279
|||||

| | | | |
|----|------|---|------|
| QY | 2521 | gtctgaagttacaattcttggagcaaaagatccctccatccagggagaatlyggagggccaag | 2580 |
| Db | 2280 | ggttaattagttacattttttggagcaaaagatctcccttccaggaagattggagggccaaag | 2339 |
| QY | 2581 | gaatccctatgtttaagaatttactctccctcagaatgaagtgtataaataatgaagaagac | 2640 |
| Db | 2340 | gaattcctttatgtttttaaattttacttcttccacacagaaatgtatataaataatgaagaagac | 2399 |
| QY | 2641 | aaaaacagtcagaanaaacttggaaoccaaelygaacccagacttcaatttactctctgt | 2700 |
| Db | 2400 | aaaaacagtcamaaaaacttttgaaacccaataatggaaacagactttcatatttctcctgt | 2459 |
| QY | 2701 | ccacccagaagaattccgttgaacgaatgtcgtgaataatcacttgggtcaagctaaagt | 2760 |
| Db | 2460 | tcaccacaaagaaatttccgggaacgaattgttcgaaatccactttgggaatcaagctcgaat | 2519 |
| QY | 2761 | tcgaagaagaagagagcaattctctaagagatatttaattgaatttgaacaacagcttgtc | 2820 |
| Db | 2520 | tcagagaggaagaaagtcatttcttagagagatttttaattgaatttgaaacagctctgttt | 2579 |
| QY | 2821 | agaatgtagccgcactgtgtataagctgcgaaccocatgtatgtctctcaatggcactccc | 2880 |
| Db | 2580 | agattgatgaaccacactcgtgtacaaacttcagaccacatgattgtctcattgtccactccc | 2639 |
| QY | 2881 | tcgcctctcccccatactgtcccccggggagagctccatgtgagagagcccaagcgcaagct | 2940 |
| Db | 2640 | tcacaccttctccattatattgtccacggagacagcttcattgagagagcccaacagcaggttt | 2699 |
| QY | 2941 | gcacaaagtctgaagaagaataagtgtacagtgtaagtgtcgtactacgaactgcgaagtgcgt | 3000 |
| Db | 2700 | gcacaaagctcgaagaagaataagtgacagtgtaagtgtcgtactacgaactgcgaagtgcgt | 2759 |
| QY | 3001 | gggaagtatgttcagattatcgcacaacaaatggccggaatcttcaagctcccaagctctgtccgt | 3060 |
| Db | 2760 | gggagtgctgttcacattttaccgacacagatggccggatctttcaaaacctccacattatcagt | 2819 |
| QY | 3061 | ggccgaacaagtctgtcatcaataatcttgtcaaccaatcaggtgtctccatccatcgagtga | 3120 |
| Db | 2820 | gccagaaacagtgatgtcatcaataatcttgtcaaccaatcaggtgtctccatccatcgagtga | 2879 |
| QY | 3121 | tgcttataggaagaacaaagtgtcatgtctgtcgtatgtgcccctccctcttaagaagatgtgga | 3180 |
| Db | 2880 | tggttataggaagaagctatggtatgtatggtgctcgtatgtcctctctcaaaagaaatgtgga | 2939 |
| QY | 3181 | acaggggacacccgaaggagacagtgctactgtgccaattacaacaacatltagccgaatgtga | 3240 |
| Db | 2940 | acaggggctttggagggacacgtgtctactggccattttacattatgacggagatggagacag | 2999 |
| QY | 3241 | aaaccgtgtcatgtatgtgaaccaactactcttccagatagaagacagggatgtgtgaagagaga | 3300 |
| Db | 3000 | aacacgctgtcatgtgacgcacccatttactcttcagaaagagacacgggaattgtgaagacagaga | 3059 |
| QY | 3301 | tagaagaacatatcacagatccagatccaaacgaacaaagcgctctccctctagagcggaacac | 3360 |
| Db | 3060 | tagaagaaccatatctacaaatccagatccaaacgaacaaagcgctctccctctagagcggaacac | 3119 |
| QY | 3361 | caaccgcgtccagatccctcttgaaagctcgtatacaaaactatagagtgtcatgtgcttcaat | 3420 |
| Db | 3120 | caaccgcgtccagatccctcttgaaagcgtatgaaacgagaaacccctcatgagagcttcatatt | 3179 |
| QY | 3421 | aatgacttggaagaatctgcccctccctctcaacccgacctatcagaggtctcaaccctgtacagg | 3480 |
| Db | 3180 | aatgacttggaagaatctgcccctccctctcaacccgacctatcagaggtctcaaccctgtacagg | 3239 |
| QY | 3481 | gtctgtccagaaagaagcccatcaagtactccgggaacaggaacgaagggtcgacagcttcc | 3540 |
| Db | 3240 | ctctgtccagaaagacactcgtcctcaagatctccgtttacacagagacgaagggccggacatttcc | 3299 |
| QY | 3541 | aaagcttccacaaagggaagaactttggagagaagtgctatgtgatatagaggagaagaatcgy | 3600 |
| Db | 3300 | taagcttccacaaagggaagacactctggaag----- | 332 |

| | | | |
|----|------|--|------|
| QY | 3601 | ccaatgaaacttaacaatatacaacaggtagccgagtcagaccoccgactgtgacgaaga | 3660 |
| Db | 3330 | ----- | 3329 |
| QY | 3661 | ttaccattcgaaaglatatcgcttcaggatgtagatccacatagagggtcagatactgtttccgac | 3720 |
| Db | 3330 | ----- | 3329 |
| QY | 3721 | taaatcttcgacagtgatgtlaagtgatgtatctgcggttccaaggacagtagtgcttc | 3780 |
| Db | 3330 | ----- | 3329 |
| QY | 3781 | tcggttcacgacgacaaagctacatgctccgtcccaatccagagcgcccgaggaggaacagga | 3840 |
| Db | 3330 | ----- | 3329 |
| QY | 3841 | aatcagtgctcttaccatcccaaatgtcaaaacagacagatggcggtgtctcggtgaagaactt | 3900 |
| Db | 3330 | ----- | 3329 |
| QY | 3901 | gaaccaaaagcaccagcatcagtgtagacatgtgtctcaactgtgagaagaatgtacgycagcca | 3960 |
| Db | 3330 | ----- | 3329 |
| QY | 3961 | gtccgacacgtcagtgtagcgcccttggttaccagtgtagcaagaacgycgatactgacatgg | 4020 |
| Db | 3330 | ----- | 3329 |
| QY | 4021 | ggccaataatgttagctattgttggctctccacggaanaagtcgcagtgctctccacaactcag | 4080 |
| Db | 3330 | ----- | 3329 |
| QY | 4081 | ccaacccgaaggaaggaagttaaagaactacgagacactgttccaagaaagcaccggaacgag | 4140 |
| Db | 3330 | -----AAGAGAGGTTAAAAAGCTTACGCACACTGTGTCCAGAGAAGCAGSAGACGG | 3380 |
| QY | 4141 | gctagcagtgtagatgtaggaacatgtgacccgcagccagccgagcgggaattccaagatg | 4200 |
| Db | 3381 | GCTAGCAGTGGAGATGAGGAACCTGGATGACAAGACAGGACCGGGAATCCACGGATGg | 3440 |
| QY | 4201 | cagcatgaaacagctatagcttcggaaaggaatctgtactctccctgggtccgccttgacctc | 4260 |
| Db | 3441 | CAGCATTGACAGTTACAGTCCAGAAAGCAAAATTGTGATCTTTCTCGGGGTCCGCTGGCCTC | 3500 |
| QY | 4261 | tgacacgcaggttcagtgatcttccctgtgtagtgcctggcgccctgcacgtatgtgacgcga | 4320 |
| Db | 3501 | TGACAGCCAGTTACACGATTTCTTGATGTCTGGGCTTGCTGCTCAGCTAGTGGAGGCCA | 3560 |
| QY | 4321 | gaacctgtgctatctctgtcaatgtgtgacacttaagrtgtgaaatgatatgtataaagaagaca | 4380 |
| Db | 3561 | GACCTGTGGACTCCTCGCATGGGTGACATTTCAGGTGGGATCATGTGATAGAAAGGAGCA | 3620 |
| QY | 4381 | gctgtaggttagaataatcattccggcgcgcgccctgtgtgttaaaacccaggttcaagacact | 4440 |
| Db | 3621 | GCTGAGGTAGAGATCATTCGGGCTCGAGGCTTGTAGTAAACCAAGGTTCCAAAGACACT | 3680 |
| QY | 4441 | gccacgacccgtatgttcaaatgttatctgtttagaacacgagctgtgatatgccaanaagaa | 4500 |
| Db | 3681 | GCCACACCATATGTTAAAGGTGACTGTTAGCAACGGAAGTGTGCATACCCAAAAGAA | 3740 |
| QY | 4501 | aaccgaagtgtgagaaagaccccttgtagccctctgtaccagcagctctgtctctcgagga | 4560 |
| Db | 3741 | AACCAAGTGGGAGAAAGCGCTGGAGCCCTGTATCACACACTGTTATCTTTCGAGA | 3800 |
| QY | 4561 | gagcccccagggaggggtgttaeaagatcatgtctgtggagattatgtgtgatgtatga | 4620 |
| Db | 3801 | GAGCCCCCAGGGGAAGGTGTTAACATCATTTGCTGGGGAGATTATGCCCCGATGGATCA | 3860 |
| QY | 4621 | caaatccttatgtgaggtgtgcccagatactcttagatgaaacttgaaatccaaatcgtgt | 4680 |
| Db | 3861 | CAATCTCTTATGTGAGTGGCCACAGATACTTAGATGAAGTGGAACTATCCAAACATGGT | 3920 |
| QY | 4681 | gattgtagtgttcaaaccttccctctccctccctccctgtgtgatccaacttcgacactct | 4740 |

```

|||||
Db 3921 AATGGATGTTCAAACTCTCCCTCCCTCCCTAGTAGATCCAACTGGCACCTCT 3980
|||||
QY 4741 gacaaagaagcttcccaatgctctctggaagtcttaccgaccttctactctgttc 4800
|||||
Db 3981 GACAAAGAAGACTTCCCATCTCTCTGTGMAAGTTCTACTGGACCTTCTACTCTGTTTC 4040
|||||
QY 4801 atagaacataaactctgttcacaaacagcagatataaaacagaagaacagca 4860
|||||
Db 4041 ATGACACTATTAACCTGTTGTTCACACACACACAGCCTTACAAAACAGAGAACACCACC 4100
|||||
QY 4861 caagctgaagacccctgtgaacaactgcatgctgtagtctgtcttacaagaccagctcta 4920
|||||
Db 4101 CAGGTGG-GGAACCTGGTCCCACTGCATGCTGATGTTGTGTCTCGGAGACCCACGCTTA 4159
|||||
QY 4921 gggataccagaagcagtcgtctgtcttcagaagaag----tcgtaacatctgtccctagca 4976
|||||
Db 4160 GGGATTAAGAGCCATCTGTGCTCTCAGAGAAAGCATCCGTGCACAGTGTGCCCTAGCA 4219
|||||
QY 4977 aagg 4980
|||||
Db 4220 AAGG 4223

RESULT 7
AF199323
LOCUS AF199323 5031 bp mRNA ROD 04-JUL-2000
DEFINITION Rattus norvegicus Rim2-2a (Rim2-2a) mRNA, complete cds.
ACCESSION AF199323
VERSION AF199323.1 GI:8925859
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 5031)
Wang Y., Sugita, S. and Sudhof, T.C.
The RIM/NIM Family of neuronal C2 domain proteins. INTERACTIONS
WITH TAB3 AND A NEW CLASS OF SRC HOMOLOG 3 DOMAIN PROTEINS
J. Biol. Chem. 275 (26), 20033-20044 (2000)
2 (bases 1 to 5031)
Wang Y. and Sudhof, T.C.
Direct Submission
Submitted (27-Oct-1999) Center for Basic Neuroscience, The
University of Texas Southwestern Medical Center, 6000 Harry Hines
Blvd., Dallas, TX 75235-9111, USA
FEATURES
SOURCE Location/Qualifiers
1. 5031
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
1. 5031
/gene="Rim2-2a"
52. 4110
/gene="Rim2-2a"
/codon_start=1
/product="Rim2-2a"
/protein_id="AAF81645.1"
/db_xref="GI:8925860"
/translation="MSAPLPGPRPAPPAASOPPPPEMPLSHLTTEERKLIQVM
DROKKEERKESVSLKILHOEFMYKEQYKKMEFEESGQOEOGADATCCICHTKPAD
GCCINCSCTCFKARCGGRVSLRNSKNVWVCNLKROEILITKSAWPIYNSGNPQ
QPDKALGLRSEEAPOEKAKLHEOTPOGPPGDSVAVERGRAHGLTRDSINSG
SGMKHQIASDMPSDNRKSPSVSRDNRKRDQSEEREYSQYVPSDTMPSDPAADR
SGREPOFYEEDPHLNYRDSNRGRHRSKEYIVDEEDVESDPEYEROREEEOAAR
SDPLARVPPOPEEOMRIHAEVSRAHERHSVSLANLEDSRLSLCHMDPS
ROBSVRRPAMENORSYSMERTRAOOSYPORTTNSPPTPRSPIDIDPELRR
ADSLRKQHHLDPSAVKTKRKMETMLRNDLSLSQSSVPRPPRPKSKGGMK
QVSLSSSEELASPEYTSQDVEITESVGKGEYSLWEHSAHSSSEASPSLHP
VTWQPSKQDRLIGRLILNLRKLDGSPADSGAMGLKVVGKMTSSGRICAPITVK
KGSIADEVGRLPGEVLEWNGRLQAGTFFEVNIIIESKEPEQVELVSRPISGMFV
RIPDSTHAQLESSSSSFESOKMDRPSISVTSFMSPMGLRDVQFLSGQLSIKLMFQV

```

```

GHOILVITLGAKLDPREDGRPNRYKLYLEPDRSDKNKRRTKTVKLTLEKMNQF
IYSPVHRPERFERMLEITLMPQARVARESESEFILLELEFALLDEDEHNVKLOHD
VSSLP.PHPSPYMRROLHGESPMRRLRSRIRISNDSYDPCDEGVGVYRHDCR
DLOSTSLVPEQVWSSNHCSPSGSEHPHADVIGRTSRMSBVSYPPOANVOGILGRTRAT
GHYNTISRMDRHWDDHYSSERDSHFLTLPSRRHRYSEHHNRGRDCEADRDQYH
RSRSTEDRPLERTTRSRSSERADNTIMRSNPSLMTGRSAPPSPALSHSRPTGSVO
TSPSTPYTGRGRQOLPOLPKGTLERGGKLRSTVOSTGLAEVRNMNTROAS
RESTDGSNNSYSSGNIIFPGVRLASDSOFSLDGLGPAOLVGOTLATPAMGDIOV
GMMDKQOLPEYITIRARLVYKPGSKTLPAPYKYLVLNDNGCIKKTKKVAKLTLEP
LYOOLSPRESFQGVLOIYWGDIGRDHSFMYVQAILLELELSNNVIGMFKLEP
PSSLVDPLTLATRRASQSLSSSPSYSR5"
BASE COUNT 1503 a 1166 c 1246 g 1116 t
ORIGIN
Query Match 57.4%, Score 2860.4; DB 94; Length 5031;
Best Local Similarity 78.0%; Pred. No. 0;
Matches 3938; Conservative 0; Mismatches 261; Indels 851; Gaps 7;
QY 1 gcttcctagggtggtgctgcgcgcgcaacatgctgctgcgtggcccgaggcg 60
|||||
Db 21 GCTTCCTAGGGGTGGTGGCTCCACCAACATGTGGCTCCACTCGGGCCCGGGCG 80
|||||
QY 61 ccggtcccaaccggcgagctctcaacctctccgcagcccgagatgcgagacctca 120
|||||
Db 81 CCCGCTCCACACCCGGGGGCTCTAGCTCTCCGAGCCCGAGATGCCGACCTCAG 140
|||||
QY 121 ccaactcaaggaagagaggaagaatacctcgtgctgtcatgagtctgcagaaga 180
|||||
Db 141 CCACCTCAGGAAGAGAGAGGAGAAATCATCCAGGCTGTATGATGTCAGCAAGAA 200
|||||
QY 181 agaggaagagagagcagtcgctgcgtcaagaatcaagaagaacaagcacaccag 240
|||||
Db 201 AGAGAGAAAGAGAGAGCTCCGTCTCAA----- 227
|||||
QY 241 gtgttcctttagtggatcactgaactggtlaataacgttctgcaagcccgaca 300
|||||
Db 228 ----- 227
|||||
QY 301 acaaccaatgagaagagcccccagaacaagctgcagccaacaattgaaatgata 360
|||||
Db 228 -----AAAGCTGATCAACAAATTGAATGATAAGGA 260
|||||
QY 361 gcaagtcagaagatggagagaagaatcgcaagcagcagaagcagaaggtgatgccc 420
|||||
Db 261 GCAAGTCAAGAAATGGGAGAGAAATCACACACGACGCAAGACAAAGGCCAGCCC 320
|||||
QY 421 gacctgtgcatctgcacaagacaatttcagatgtagatggcgcataatgcttcta 480
|||||
Db 321 GACCTGCGGCATCTGCCACACACAAATTTGCAGATGGATGGCGCATTAAGTTCAT 380
|||||
QY 481 ttgccaaccaagctctgctcgatgtagagtgtagtgcttcaagctcaacaag 540
|||||
Db 381 TTGCCAAACCAAGTCTGTGCTGTGTGAGAGTGCAGTGTCTATAGCTCAACAAG 440
|||||
QY 541 tatgtgggtgtgtaattgtgtccgaaacaacaagaatctctactaataacagagat 600
|||||
Db 441 TATGTGGGTGTAAATTGTGTGCCGAAACACACAAATCTCTACTPAAGTGGCGGCT 500
|||||
QY 601 gtttataataatgggtctcaacacacatgcaaaccttgatccaagaagtctcgaag 660
|||||
Db 501 GTTTACATATATGGGTCTAATACACCGCAGACCTGATCAAAAGGCTTTTGAGGCT 560
|||||
QY 661 tcgaatagagaagccctcagaagaagaagaactacacagcagcagccagttcca 720
|||||
Db 561 TCGAGTAGAGCAAGCCCTCAGAGAAAGCAAACTGCATGACAGCAGCAGTTCGA 620
|||||
QY 721 aggaagcccaagtgactatacagctcctcagttgagaagcagcggtcactagct 780
|||||
Db 621 GGGACCCCGGTGACTCTCATCTCAGTTCAGTGAAGAGGTGAGCTCATGAGGCTC 680
|||||
QY 781 aagaagataataataaataatgataagagtgaaagcacaatatgacagatgac 840
|||||

```

Db 681 AAGACGATCTATTAGAAATGGCTCAGGAATGAAGCAACCAGATGCCAGTGACATGCC 740
 QY 841 ttccagacgaaacgaagccatcagtcagtcagggatccaatccgaatcagcagcaag 900
 Db 741 TTCAACACAAAAAGAGTCCTGACGTGTCAGGGATCAAAATCGAAGATACGACCAAG 800
 QY 901 tgaagaagaagagactacacagiatgttccttcagatggtacaaatgccaaatctcc 960
 Db 801 TGAAGAAAGAGGAGATATTCACAGATATGTTCTTCAGATAGCACAATGGCTAGATCTCC 860
 QY 961 ttcgagattatgcttagaagacatctcagcgttgagccccaatttataagaactgttca 1020
 Db 861 ATCAGATTTGCTGATAGAGATCTCAACGTGAACCTCAATTTTATGAGAACTGTGATCA 920
 QY 1021 tttaaattcaaggatctctacaggaagagccatagacattccaagaaglatatgttga 1080
 Db 921 TTTAAATTTATAGGATTTCTTAACAGAGAGCCATAGACATCCAAAGAGTATATTGATGA 980
 QY 1081 tgaatgaagatgtgagagacagagatgaatatgaagaacaaagagagagagagaaatga 1140
 Db 981 CGACGAAGATGTGAGAGCAGAGATGAATGAATAAGAAAGAGAGAGAGAGATACCA 1040
 QY 1141 ggcagcctaaagaagatgaatccaatctgcgcggtatcccgtaaagccacaaccctaca 1200
 Db 1041 GGCACGCTACAGAAAGTATCCAAATTTGGCCGGTATCCGGTAAAGCCACACCTATGA 1100
 QY 1201 agacaacatgcgacacagcgtgagatgttccaagggcacgacatgagagagacagatga 1260
 Db 1101 AGAGCAAAATCGGATCCACGCTGAAGTGTCCGGGACAGACAGAGAAAGCACAAGTGA 1160
 QY 1261 tgttctcttggcaacgctgacatagaaagatccaagatctctcgtcgaagatgatag 1320
 Db 1161 TGTGTTCTTTGGCAAAATCTGTAAGATGAAGATTCAGGATTTCTGCTAGAGATGATAG 1220
 QY 1321 accatcaagcacaagatctgtatctgtaacgtagagctgcaatgtaaaaaaccaacgata 1380
 Db 1221 ACCATCAAGCAAAAGATCTGATCTGAACGTAGAGCTGCAATGGAATAACAGAGATCGTA 1280
 QY 1381 ttcaattgaaagaaactcgaagagctcagggagacaagttcttatccaagaagaccctcaa 1440
 Db 1281 TTCAATGGAAGAAACTCGAGAAAGCTCAGGGAGCAAAAGTTCTTATCCAAGGACCCAAA 1340
 QY 1441 tcaatagtcctccacccctcggcgagacccctataccgcttgaagaccagacatgagcg 1500
 Db 1341 TCAATAGTCTCTTACCCCTCGAGAGCCCTTATACGCTCGATAGACAGAGCTGAGGCG 1400
 QY 1501 cgcctgactccctacggaacagcaccacttagatcccaagctcgtcgttgaagaaacgaa 1560
 Db 1401 TGCCGACCTCCATACGGAACAACACCACTTAGATCCACGCTGCTGTAAAGGAAAGAA 1460
 QY 1561 gcgagaaaaaatggaacacatgtttaagaaatgattcttgaagttcagaccagtcgagtc 1620
 Db 1461 GCGAGAAAAAATGGAACAAATGTGAGATGATTTGTGAGTTGACACCAAGTCTGAGTC 1520
 QY 1621 agtggagcgccccccacccaagcctcatataatccaagaagaagagatlaaaatgtcgcaag 1680
 Db 1521 AGTGAGGCGCCGCCCAACCAAGGCTCATTAATCCAGAAAGAGAGGATTAATGCGCCAGGT 1580
 QY 1681 ttcaattgacagcctcggagagagagcttgacatccacacactgagtatatacaagctgttga 1740
 Db 1581 TTCACTTGAGCAGCTCTAGAGAGAGAGTGTGGCATTCACGCTCGAATATTAACAAGCTGTGATGA 1640
 QY 1741 tctgagactggaagagagagatgttgtagtgaagaaggagacagtcacaagaaggaaaaaaga 1800
 Db 1641 CGTGAAGATTGAAAAGGGAAGCGTAGGTGAGAAAGGAGAA 1681
 QY 1801 aactagtgagcagggaggtttgtcgatcttaacacacaggtcttgagagacaaaagaagaag 1860
 Db 1682 ----- 1681
 QY 1861 gatgtactatgttgcacactcttggaaagagatttggaaatggtctgagcctcagattaa 1920
 Db 1682 -----TGAGTACAGCTGATTGAGAGCATGCGCTTTG 1712

QY 1921 ggaacttggggtagataactcgttagtagcaacaacccttaacgaagagcatagccatagta 1980
 Db 1713 G-----CATAGCAAGAGAGCATCCCAATGTC 1739
 QY 1981 taagcaccttgaacttgcagacccatcccaagaatgagatcgccctaaatvtgttgtatltt 2040
 Db 1740 TTTGCACCTGTGACCTGGACGACCATCAAAAGATGAGATGCGCTTAATTTGCTGATTTT 1799
 QY 2041 attaaatagcgttttaagaatgagagtgtagtctgagactcagactcaggaacaaatgtcgagct 2100
 Db 1800 AATTAAATTAAGCGTTTAAAGATGAGGTGATCTCGAGATTCCAGAGACATGCTGGGCTT 1859
 QY 2101 aaaggttgaagagagaagaatgaactgaatcagtcagacttgttgcattatlaacaaagt 2160
 Db 1860 AAGGTGTAGAGGAGAAAGATGATGATAGAGTGCAGCTTGTGCAATTTATACCAAGT 1919
 QY 2161 aaaaaaagaagtttagctgtatcctgttagaactccttaagccagtgatgaagcttggga 2220
 Db 1920 GAAAAAGGAAGTTTATGCTGATCTGTAGGACATCTTAAGCCAGTGTGATGAGTCTTGA 1979
 QY 2221 atggaaatggagagctatctgaagagagccacatttgaggaagtttacaacattatctaga 2280
 Db 1980 ATGGAATGGAAGCCTATTGCAAGAGCCACATTTGAGAGAACTTTACAACATTTATCTGGA 2039
 QY 2281 atccaaactgaaacacaagttgagcttgttltccaagccaaatggagatatctccag 2340
 Db 2040 ATCCAAAGCCTGAAACCAACAGTTGAGCTTGTGTTCAAGCGGATTTGAGATATGCCTAG 2099
 QY 2341 aatacctgataagcaagcagac 2400
 Db 2100 AATAACCTGATAGCACCTATGACACACTGGAATCCAGATTCTAGCTCATTTGAATCCAAA 2159
 QY 2401 aatggacgctccctctatccgttatccctacacacacacacacacacacacacacacacac 2460
 Db 2160 AATGGATCGTCCTCTATATGCTGATACCTCTCCATGATGCTGATGCTGATGATGATG 2219
 QY 2461 cccgagcttctatcttgcagacagcttccaataaactatggttgcagaggttgcacaca 2520
 Db 2220 CCCACAGTTCTTATCTGACAGCTTTCAATTAATGATGTTGACAGAGTGTGACCA 2279
 QY 2521 gttcagatgtacaatttggagagcaagatctccctcccaagagagatgagagccaaag 2580
 Db 2280 GTTATATGTTCAATTTTGGAGCAAAAGATCTCCCTTCACAGGAAAGATGGAGCGCAAG 2339
 QY 2581 gaatccatgtttaaagattacttccctcccaagaagaagtgataaataaagaagaagaac 2640
 Db 2340 GAATCCTTATGTTAAATTTACTTCTTCCAGACAGAAAGTATTAATAACAGAGAAAGAAC 2399
 QY 2641 aaaaacagtcagaagaacttggaaacccaatggaaacagacttccattatctcctgt 2700
 Db 2400 AAAAACGTTCAAGAAACTTTGGAAACCCAAATGGAACCAAGACTTTTCAATTTATCTCTGT 2459
 QY 2701 ccacgagagagatccgtgaaacgaatgctggaatlaaccccttggatcaagctagagt 2760
 Db 2460 TCACCGAAGAGATTTCCGGAGACGAATGCTGAATACACCTTTGGAGATCAACCTGAGT 2519
 QY 2761 tgcagaagaagaagcgaattcttaggagagattttaattgaatttgaacaacagcttgc 2820
 Db 2520 TCGAGAGGAAGAAAGATTTCTTAGGAGAGATTTTAATTTGAATTTGAAACACCTCTGTT 2579
 QY 2821 agatgtagccgcagcttggtatagaagtgagagaccatgtagtctccctacatgcacatccc 2880
 Db 2580 AGATGATGAACCACTGTGTACAACTTCAGACCCATGATGTCTCCTCATTTGCCACATCCC 2639
 QY 2881 tgcgcttcccatatctcccgagagcagctccatctgagagagcccaacgcgcagagct 2940
 Db 2640 TCACCTTTCATATATGCCACGAGACAGCTGATGAGAGAGCCACACGCAAGTT 2699
 QY 2941 gcaaaagtgcgaagagaataagtagacagtgagatgctcgtacagactcagagcagagagag 3000
 Db 2700 GCAAAAGCTGAAGAGAAATAGTAGACAGCGAAGTATCGACTATGACTGCGAGGAGCGCGT 2759

| | | | |
|----|------|---|------|
| QY | 3001 | gggagtagtgcataatcgcacaacatggcgcgagctcttcaagctcacgtgtgtcgt | 3060 |
| Db | 2760 | gggagtcgtcgcagattaccgcacacgatggccctgattctcaaggctccattatgagt | 2819 |
| QY | 3061 | gccaaacaagcatgctcaatacatgtgtccacatcaaggtctccatcgcagtaga | 3120 |
| Db | 2820 | gccaaacagatcatgtctatcnaaacattgtctaccatcagggctctctcatcgagtaga | 2879 |
| QY | 3121 | tgctatagaaagacaagatctcgtgcgcttagtgccctccctccctcaaaagaa | 3180 |
| Db | 2880 | tgctatagaaagacagtagtctatgtgtgcgcttactgtgtccctcaaaagaa | 2939 |
| QY | 3181 | acaagggcacggaaggacaagtcgtctactcgtgcacatacaacaattagccgaatgtag | 3240 |
| Db | 2940 | acagggcgcttcagaggacacgtgctactgcccattcaatatactataggccgagtagac | 2999 |
| QY | 3241 | acacccgtgtcaatgtagtaccactactcttcagataagac | 3280 |
| Db | 3000 | acacccgtgtcaatgtagtaccactactcttcagataagac | 3059 |
| QY | 3281 | -----aggatgtgaagc | 3294 |
| Db | 3060 | tcgcctctgcacacaggcacaacagtagatcatccacacaggagatggaaggattgtgaac | 3119 |
| QY | 3295 | agcagatagacagccatalcacagatlcacagatacaagaacaacgycctctcctagaagc | 3354 |
| Db | 3120 | agcagatagacagccattatcacagatccacagatcaacagaaacaggcctctcctagaagc | 3179 |
| QY | 3355 | gacacacacccgctccagatctctctgaagctcctgtatataaacctcatgtagttagcc | 3414 |
| Db | 3180 | gaccacacacccgctccagatctctctgaagctcctgtatataaacctcatgtagttagcc | 3239 |
| QY | 3415 | ttccatatagtatgtagaattcgtgcctccctccctcaactgtcattacgagtcacacctgc | 3474 |
| Db | 3240 | ttccatatagtatgtagaattcgtgcctccctccctcaactgtcattacgagtcacacctgc | 3299 |
| QY | 3475 | taccggtctgtccagacaagcccatcaagtactccgvgaaacagcagaaaggccgaca | 3534 |
| Db | 3300 | cactggctctgtgccagaccagtcgcctcaagtagtctcccggttaacagagacgaaggccgaca | 3359 |
| QY | 3535 | gcttccacagcttccacacaaaggaaacatttgagagagaagtgtcattgatalaagagagag | 3594 |
| Db | 3360 | gcttccacagcttccacacaaaggaaacatttgagagagaagtgtcattgatalaagagagag | 3395 |
| QY | 3595 | aaatcgccaatgaaacttaacaataaacaacagtagccgagatcagaccocagatcga | 3654 |
| Db | 3396 | ----- | 3395 |
| QY | 3655 | gcaagatcacatcgcgaaglatcgtccagatlggatalccacatagaggggcagaatactgt | 3714 |
| Db | 3396 | ----- | 3395 |
| QY | 3715 | ttccactaaatcctcgcgaagcgtgataagtgtatctcgcggttccaagactagtag | 3774 |
| Db | 3396 | ----- | 3395 |
| QY | 3775 | tgctctcgtttccagcagcaaaagctacatgtccgtcccaatcagagcggccgagaggaa | 3834 |
| Db | 3396 | ----- | 3395 |
| QY | 3835 | caggaaatcagtgctcttaccatccaaatgcaaaacagacaagatgggcgtgtcggggaa | 3894 |
| Db | 3396 | ----- | 3395 |
| QY | 3895 | gaacttgacaaaagcacacagcatcagtgtagacatgtgtctcactgtagaagaatgacgg | 3954 |
| Db | 3396 | ----- | 3395 |
| QY | 3955 | caggcagttccgacatcgcaggtggcgccctgggtaccaggtgccaagaagcgcgactag | 4014 |
| Db | 3396 | ----- | 3395 |
| QY | 4015 | catggggccaatgtagtactatgttgtgtctcacggaaaagtgcagtgctctca | 4074 |

[illegible]

| | | | |
|------------|-------------------------------------|---------|---------------|
| RESULT | 8 | | |
| AF199326 | | | |
| LOCUS | AF199326 | 5031 bp | mrna |
| DEFINITION | Rattus norvegicus RIM2-3B (Rim2-3B) | mrna | 04-JUL-2000 |
| ACCESSION | AF199326 | | complete cds. |

```

VERSION      AF199326.1  GI:8925865
KEYWORDS
SOURCE       Norway rat.
ORGANISM     Rattus norvegicus
              Chordata: Craniata: Vertebrata: Euteleostomi:
              Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae:
              Rattus.
REFERENCE    1 (bases 1 to 5031)
AUTHORS      Wang, Y., Sugita, S. and Sudhof, T. C.
TITLE        The RIM/NIM family of neuronal C2 domain proteins. INTERACTIONS
              WITH RAB3 AND A NEW CLASS OF SRC HOMOLOGY 3 DOMAIN PROTEINS
JOURNAL      J. Biol. Chem. 275 (26), 20033-20044 (2000)
MEDLINE      20347919
REFERENCE    2 (bases 1 to 5031)
AUTHORS      Wang, Y. and Sudhof, T. C.
TITLE        Direct Submission
JOURNAL      Submitted (27-OCT-1999) Center for Basic Neuroscience, The
              University of Texas Southwestern Medical Center, 6000 Harry Hines
              Blvd., Dallas, TX 75235-9111, USA
FEATURES
source       1. 5031
              /organism="Rattus norvegicus"
              /db_xref="taxon:10116"
              1. 5031
              /gene="Rim2-3b"
              52. 4110
              /gene="Rim2-3b"
              /codon_start=1
              /product="Rim2-3b"
              /protein_id="AA01648.1"
              /db_xref="GI:8925866"
              /translation="MSAVLGPGRPAPPPAASQPPPEPMDLSLTFEERKTIQAVM
              DROKKEEKEDSVLKLHQDFEMKKEYKKGEESQOQEQGDAVTCCTCKTKRAD
              CGCHNCGYCOTFCARGCGRYSLRSNKVMVNCILCKQOELLTKSGAMYNQSGNTPQ
              OPDOKALRGLESEAPQEKAKLHQTOFCQCPGSDSYPAVGRAGHGLTRDSDIKNG
              SGKMHQIASMDPSRKSRSYSDQNRREYDSEESQYSPSTORMRESEYQAYR
              RSQREPPFYEPDHLNYSRNSRGRHRSKEYIVDEEDVESREYREYRESEYQAYR
              SDPMIARYPVKQPYEEQMRIHAERYRAHERHSDVLSANMELDSRISLIMRDS
              RQVRSVSRRAAMENQRTSMETRAQCSSTPQRTTNSHPTPPRSPILPIRPELR
              ADSLRKHHLDPSSSAVKTREKMETMLRNSLSDQSSVPPPPRPHKSKGKMR
              QVSLSSSEELASTPEYSCDVELESSEVSGKDMEXSWLHASMSSEASPSLHP
              VTWQPSDGRDLIGRLILNKLKDGSPVDSGAMGLKVGKMTSGSLCAEITLVK
              KSLADPVAHDPGDEVLEMGRLLOGATFEERYVILLESKEPOVELLYVSRIGDM
              RLPSTADYAOLESSSSSPESOKMDRPSISVTSFMSMGMLDVPOLSGOLSIKLMPKY
              GHQILVITLGAKLDPSEDFGRPNRYKTYFIPDSDKRKRTKYKTKLEPMKQTF
              IYSPVHREFRERMLEITLMDQAVRSEESFLEGLLELTALEDDEPHWIKLOTHD
              VSSILPIHPSPYMPRLHGESPRRLQSKRISDSVSDVDCEDVGVVSDYRHQK
              DLOSSTLSVPOVSNHSPSGSPRHVLIQRTSRWSPSPVPPORNNVDGLRQAT
              GHYNTISMDRHYVDHYSERDSHFLLPRSRHQTSEHHRRDGDCEADARQYH
              RSRSTEQPRLIERTTSSRSRADTNLMRSMPLMTGRSAPSPALSHPSTGSAVO
              TSPSTPYTCGRKRLPOLPPKGTLERGGCKLRSTVORSTETGLAVEMRNMPTROAS
              RESIDGSMNYSVSEGNLIFPGVRLASDSQSPFLDGLGROLVGRQTLATPAAGDIOY
              GMMDKQGLEVEILIRAGLIVKPGSKTTPAPYKVTLLNGVCTIAKKTKVAKTLEP
              LYQOLISFEESPOGKVLQIIVWDYGRMDHKSMFYAQLLDLELSNMVITWFKLEP
              PSSLYDPTLAPLTRASOSLESSTGSPYSRS"
BASE COUNT   1503 a 1166 c 1246 g 1116 t
ORIGIN
Query Match      57.4%; Score 2860.4; DB 94; Length 5031;
Best Local Similarity 78.0%; Pred. No. 0;
Matches 3938; Conservative 0; Mismatches 261; Indels 851; Gaps 7;

```

```

Db 141 CCACCTCAGGAAGAGAGAGAAATCATCCAGGCTGTCTGATCGTACAGAGAAAGA 200
Oy 181 agaggaagagagcagtcgctgtcctaagatccaagaagaacacaaagcacaccgacaca 240
Db 201 AGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 227
Oy 241 gtagtcccttagtggagatcaactgtaataacgltcgtcagccccagcaaaa 300
Db 228 ----- 227
Oy 301 acaaccatgagaagagagccccaacaaagctgcacacaaattgaatgataaga 360
Db 228 -----AAAGCTCAGCAACAAATTTGAAATGTAAAGGA 260
Oy 361 gcaagtcagaagatggaagaagatcgagcagcagcaagaagcagaaggtgtagtccc 420
Db 261 GCAAGTCAGAAATGGAGAGAGATCAACACAGCAGCAAGAGAGAGAGAGAGAGAGAGAG 320
Oy 421 gacctgagcactgcacacaagaacaaattgcagatgtagtcgagcacaattgtccta 480
Db 321 GACCTGGCGCATCGCCACAAAGCAAAATTTGCAGATGATGCGCGCCATTAAGTTCATA 380
Oy 481 ttgcacaaccaaagtcctgctcgtatgtggaagtcgaagtcgtcttaagctaaacaagt 540
Db 381 TTGCCAAACCAAGTTCTGTGCTCTGTGGAGTGTGATTAACGCTCAACAAAGGT 440
Oy 541 tagtgagtgtagtaattgtgccgaacaacaagaatccctactaaatcagagcagtcg 600
Db 441 TATGTGGTGTATATTGTGCTCCGAAACCAACAGAAATTCCTACTAAGTGGCGCGTG 500
Oy 601 gtttataatagtggtcttaacacacctgcagcaacctgatalcaaaagttcctcagaggt 660
Db 501 GTTTTAAATATGTGGGTCTAATATACACCCAGAGCCTATCAAAAGGCTCTTCAGAGGCT 560
Oy 661 tcgaatgagaagagccctcctcaggaagaagaacaaacaaacagcagagcccaagttcca 720
Db 561 TCGAGTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 620
Oy 721 aggaagccagtgtagctatcagtaactcagtcagtcagtcagtcagtcagtcagtcagtc 780
Db 621 GGGAGCCCCCGGTGACTCATGATCATGCTGCACTTGAAGAGAGAGAGAGAGAGAGAGAGAG 680
Oy 781 aagacagatcatttaaaatlgatcagagtggaagccagatctgcagtgagactgtgc 840
Db 681 AAGACAGATCTTATTAAGATGAGGCTCAGAGATGAAGCCAGAGTCCAGATACATGTC 740
Oy 841 ttgcagcaagaagaagatccatcagtggtccagggggtccaatcgaagatacgaagaag 900
Db 741 TTCACAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 800
Oy 901 tgaagaagagagagactactcaacagtagtctcctcagatggtlacaatgccaaagatctcc 960
Db 801 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 860
Oy 961 ttgcagatctgctatagacagatctcagtgtagagcctcaatttgaagaagactgtgca 1020
Db 861 ATCAGATTAATGCTGATAGAGAGATCTCAACGTGACCTCAATTTTATGAGAGAACTGATCA 920
Oy 1021 ttaaatcaagagatctcaacagagagagccatagacatctcaagaagatattgttga 1080
Db 921 TTTAAATTTAAGGATTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 980
Oy 1081 tgatgaagatgtggaagagagagatgaatatgaagaacaaagaagagagagagagagatcca 1140
Db 981 CGACGAGAGATGTGAGAGAGAGAGAGATGAATAAGCAAAAGAGAGAGAGAGAGAGAGAGAG 1040
Oy 1141 ggcagagctcaagaagatgataccaatctggccggtatcccgtaaaagccacaacctcaga 1200
Db 1041 GGCAGGCTACAGAGAGATGATCAAAATTTGGCCGCTATCCGCTAAAGCCACAACTTATGA 1100
Oy 1201 aagaacaatgctcagatccagctgaggtgtccagagcagacatgagagaagagcagatga 1260
Db 1101 AAGAGCAATATGCGGATCCAGGCTGAAGTGTCCGGGACAGACAGAGAGAGAGAGAGAGAGAG 1160

```

QY 1261 tgtttcttggcaaacgctgaaactagaagaattccaggaattctctctgaagatgtag 1320
|||||
Db 1161 tgtttcttggcaaaaTgcTGAACtAGAAcATTCCAGaATTtCTGCTAGATgATgATg 1220
QY 1321 accatcaaggcaagaatctgtla tctgaacgtlagagctgcgaatgtaaaacccaagatcgta 1380
Db 1221 ACCATCAAGGCAAAAGATGTGTATCTGAACGTAGAGCTGCATGGAAGAAACCAAGGATGTA 1280
QY 1361 ttcaatgaaagaactcgagagagcttcagaggacaagaattcttatccacaagaagctcaaa 1440
Db 1281 TTCAATGGAAGAAcATCTCGAAGCTTCAGGACAAAGTTCTTATCCAAAGGACCAcAAA 1340
QY 1441 tcaatgtctctccaccctccgagcgagcctataccgctctgataagcccaagatgagcg 1500
Db 1341 TCATAGTCTCTTACCCTCTCGAGAGGCCCTATACGCTCGATAGACCAAGCTGAGGCG 1400
QY 1501 cgtgactccctcctcagaaacagcacacttagatcccaagctcgtctgtgtaggaacga 1560
Db 1401 TGCCGACTCCCTACGGAACAAcACCATTAGATCCAGCTGTGCTAGAAAGAAAGAA 1460
QY 1561 gcgagaaaaaaatggaacacatgtttaagaatgatctcttgagttcagaaccaagtcagatc 1620
Db 1461 GCGAAGAAAAATGGAACAAcATGTGAGAAcTATCTTTGAGTTCAAGCCAGCTGTGATC 1520
QY 1621 agtggagccgcgcccaacgaaggccctataatccaaagaagaaggtlaaaatltgcgcaggt 1680
Db 1521 AGTGAGGCGCGCCCCACCAAGGCTCTATAAATCCAAAGAAAGAGATGAATGCGCCAGGT 1580
QY 1661 ttcaatgacagctcsgaagagagagcttgcacatccacactgaatatacaagctgtgtag 1740
Db 1561 TTCACTAGACACACTCAGAGAGAGAGTGGCATCCAGCGCTGACTATATCAAGCTGTGATGA 1640
QY 1741 tgtgagcttggaagcgagagatgtgtgagtgagaaaggagagcagtcacaaaggaaaaagaa 1800
Db 1641 CGTGAGAGTTTGAAACCGAGAGCGTAGTGAGAAAGGAGACA----- 1661
QY 1801 aactagtgagcagggagtttltgcgattcttaaccacaggtctcgagagacaaagaagaag 1860
Db 1682 ----- 1681
QY 1861 gatgtactatggtgagccactcttctgaagaagatltgaaatgtctgcagcctcagattaa 1920
Db 1682 -----TGGAGTACAGCTGGTTGAGAGATGGCTTTG 1712
QY 1921 ggaactctgggttagatlaactctgtagtagcaaaccttcaacgagagagatagcagtagta 1980
Db 1713 G-----CAcTAGCAGTGAAGGCATCCCAATGTC 1739
QY 1981 taagcacctgtgacctgagcagccatccacaagaatgagatgcgctaattgtgtagtattt 2040
Db 1740 TTTGACACCTGTGACCTGGACGCCATCCAAAGATGAGATCGCTTATTTGTGTATTTT 1799
QY 2041 attaaatagcgtlttaaaagatgaggagtgtaacctcgagactcagagacaaatgactgagctt 2100
Db 1800 ATTAATATAGCGTTTAAAAAGATGGAGGTGTAACCTCGAGATTCAGAGCAATGCTGGGCTT 1859
QY 2101 aaaagctgttaggaggaagaatgactgaaatcgaagtcgaacttltgcatattatccaagaat 2160
Db 1860 AAAGGTTTAGAGAGAAAGATGACTGAATCAGGTGCACATTGTGCAATTATTCACAAAGT 1919
QY 2161 aaaaaaagaagaatttagactgatactgtagaactttagaacsagatgataaagttaga 2220
Db 1920 GAAAAAGGAAGCTTAAAGCTGATGACTGTAGACATCTTAGACCAAGGTGATGAAGCTTGGGA 1979
QY 2221 atggaatggaagctatgtgcaaggagccacatttgaagaagtttaacaatatltctaga 2280
Db 1980 ATGGAATGGAAGGCTATTGCAAGGAGGCCACATTGTGAGGAAGTTTACAACATTAATTCTAGA 2039
QY 2281 atccaacactgaaacacaagatgtgagctgtgttccaaggccaattggagagattctcctag 2340
Db 2040 ATCCAAAGCTGAACCAAGTTGAGACTGTGTTCCTCAAGGCCGATTTGAGATATGCTTAG 2099
QY 2341 aatacctgatagcagcatgcaacaactggaaatccaggtcttagctcaattgtgaatccaaaa 2400
Db 2100 AATACCTGATFACACTCATGATGCACAACTGGAAATCCAGTTCAGTCAATTGAATTCAAAA 2159
QY 2401 aatggaacgtctctcctatataccgttaccctcaacccatgagtccttgcagatgtgaggaatgt 2460
Db 2160 AATGATGCTCCGCTGTATATCGGTATACCTCTCCATGAGATCCTGGCATGCTGAGGAGATG 2219
QY 2461 cccgcaagttctatcttgagcagcttccaataaactatggtlttgacaaggtltgtcaaca 2520
Db 2220 CCCACAGTTCTTATCTGGACACTTTCAAATAAAcATGATGTTGACAAAGGTTGGTCAcCA 2279
QY 2521 gtgtagtataatlttggaggaagaagatccctctcccaagggaagaatgaggaggaag 2580
Db 2280 GTTAATAGTTAAATTTTGGAGCAAGAGATCTCCCTCCAGGGAAGATGGAGGCCAAG 2339
QY 2581 gaatccttatgtlaagaatttacttctctccagaatgaatgataaataaagaagaagaac 2640
Db 2340 GAATCCTTATGTTAAATTTACTCTCTCCAGACAGAGATGATAAAAcAAGAGAAAGAAC 2399
QY 2641 aaaaaacgttcaagaanaacttggaaacccaatggaaccagacttcaatttatctctgt 2700
Db 2400 AAAAAcAGTCAAGAAAAcTTTGGAAcCCAATGGAACCAAGACTTTCATTTATTCCTGT 2459
QY 2701 ccaccgaagagaatcccgtagaacgaatgctcggaaattaccccttggagataagaatagat 2760
Db 2460 TCACCGAAGAGAAATTCGGGAACGAATGCTCGAAATCAcCCTTTGGATCAAGCTCGAGT 2519
QY 2761 tcgaagaagaagaagcgaatctcttagagaagatttlaatgaaatlggaacacagcttgc 2820
Db 2520 TCGAGAGGAAGAAAGATTTCTCTAGAGAGATTTTAATGAATTGGAACAGCTGTGTT 2579
QY 2821 agatgatagcgcgcactggtataagctgcagaccatgatatgtctctcaatltgcacatccc 2880
Db 2560 AGATATGAAACCAcACTGTGTACAAACTTCAGACCAcTGAATGTCTCCATTTGCCACATCCC 2639
QY 2881 tgcgcctcccatatctgcgcgcggagagctcactcagtgagaagcccaacgcgcagagct 2940
Db 2640 TCACCTTCTCATATATATCCACGAGACAGCTGCATGGAGAGAGCCCAcACCAcAGTT 2699
QY 2941 gcaaaagtcgaagaagaatlaagtgacagtgagtgctcgaatagcagtcgagagatgagct 3000
Db 2700 GCAAAAGTTCGAAGAGATATAGTGACAGCGAGATATCCGATATGACTGCGAGAGCGCGT 2759
QY 3001 gggagtagtgcagattatcgacacaaatggcgcgagatcttcaaaagtcgaagtgtgcgt 3060
Db 2760 GCGACTGTGTGAGATTAACCGACACGATGCGCGTGAATCTTCAAAAGCTCCCAATTATACGT 2819
QY 3061 gccagaacaagtcatagtcaatcaatcatgtgcacacatgaaggctctctcatcgaatgaga 3120
Db 2820 GCCAGAAACAGTCAATGTCTCAAAACCATTTGCTACCATCAAGGCTCTCTCATCGAGTAGA 2879
QY 3121 tgtlaagaagaagaaggltcatgtgcgcctagtgccctccctccctcaaaagaaatgtaga 3180
Db 2880 TGTATTAGAAAGAGCTAGGTCAATGTGCGCTAGTGCCTCTCCCTCAAAAGAAATGAGGA 2939
QY 3181 acaggggcaacggaggaagaacgtgtcactgtgccaattacaacacaatlagccgaatgtaga 3240
Db 2940 ACAGGGGCTTCCAGGGAGCAcGCTACTGCGCAATTCAAATACAAATTAGCCGCGATGGACAG 2999
QY 3241 aacaggtgcatagtgagtgacaactactcttcaagatagaagac----- 3280
Db 3000 ACACCGTGTCAATGAGACGACCAATTACTCTTCAAGAGAGAGAGATCAATTTTCTCACTACCC 3059
QY 3281 -----aggatlttgaagc 3294
Db 3060 TCGCTCTGCACACAGGCAAGCAAGTGAGCATCAcCAAGGAGATGGAAGGATTTGTGAGAC 3119
QY 3295 agcagatagacagcatatcatcagaatccagaatcaacagaacacggcctctcctagaagcg 3354
Db 3120 AGCAGATGATGACAGCCATATACAGATCCAGATCCAAcAGAAcAAGCGGCTTCTCTAGAGCG 3179
QY 3355 gaccacccacccgctccagatctcttgaacgttccctgataaacaactatgagtgcatgccc 3414

|||||
Db 3180 GACCACCCGCTCCAGATCCTGTAAGCTGCTATACAAACCTCATGAGTCGATGCC 3329
OY 3415 ttcaataatgactggagaatctgccccctccctcaactgacctatcgaggtcacaacctcg 3474
Db 3240 TTCAATTAATGACTGGAAGATCTGCCCTCTTACCTGCTTATCGAGTCTCACTCCG 3299
OY 3475 tacgggtctgtccagacaagcccaatcaagtactccggggaacagagcggaaggcgaca 3534
Db 3300 CACTGGCTCTGTCCAGACAGTCCTGTAAGTACTCCGTAACAGAGAGAGAGGGGCGGCA 3339
OY 3535 gcttcacagcttccacccaagaagcaatctggagagagtgctatgataagagagag 3594
Db 3360 GCTTCTCAAGCTCCACCAAGGAAGACTGMAAG----- 3395
OY 3595 aaatgcacaatlgaaaactaacaatalacaacagtagccggatcagaccccgactgga 3654
Db 3396 ----- 3395
OY 3655 gcaagattaccattcgaaagtalccgtccaggatggatccacatagaggcgcaatactgt 3714
Db 3396 ----- 3395
OY 3715 ttccactaaatccctcgagacagtagtagtgaagtgtatctgcgtttcaaggactagtag 3774
Db 3396 ----- 3395
OY 3775 tgccttcglttcagcagcacaagctacatgltccgtccaaatcagagcgccgagagaaa 3834
Db 3396 ----- 3395
OY 3835 caggaataatcagtgcttcttaacatccaaatgcaaaacagacagatggcggtgctgggaa 3894
Db 3396 ----- 3395
OY 3895 gaacttgaccaaaagcaccagcatcagtgagacatgtgtcaactgagagaagaatgacg 3954
Db 3396 ----- 3395
OY 3955 cagcagctcgcagactgcagtgaggcgccctgtggtlaacagtgcgaagaagcgcgactag 4014
Db 3396 ----- 3395
OY 4015 cattggggccaatatgtatgtgtgtctctccacgaaaaagtgcagtgctctca 4074
Db 3396 ----- 3395
OY 4075 actcagccaaaccgaaggaggtaaagaactacgagacactgttcagagaagcacgga 4134
Db 3396 -----AGGAGAGGTAAAGCTAGCAGCAGCAGCTTCAGAGAACGACAGCA 3440
OY 4135 gaccgggctagcagtgagatgaggaactgatatgcgcgcagccagccagccgggaatccac 4194
Db 3441 GACCGGGCTAGAGCTGAGATGAGAACTGGATGACAAGACAGCGCCAGCGGATCCAC 3500
OY 4195 agatgacacatgaacagctatagcttcoggaagaatctgattcttcctgggggtccgct 4254
Db 3501 GGATGGCAGCATGAACAGTTACAGCTCAGAGGAATCTGATCTTCTGGGGTCCGCT 3560
OY 4255 ggcctctgacagcagcttcaatattccctggatggcctgggcccctctcagccagttgg 4314
Db 3561 GGGCTTGACAGCAGCTTCAAGCATTTCTTGATGCTGTGGCCCTGCTCAAGCTAGTGG 3620
OY 4315 acgacagacccttgactactctgcaatgggtgacatcgaatgggtgggaatgataaaaa 4374
Db 3621 ACGCCAGACCCCTGGCGACTCTCTCAATGGGTGACATTCAGTGGGATGGATGAAGAA 3680
OY 4375 gggacagctgggaagtgaataatcctcgggcgcgcgccctgtgtgtaaaacaggttccaa 4434
Db 3681 GGGACAGCTGGAGAGTGAATCCTCGGCTCGAGGCTTGTAGTAAACCAAGGTTCCAA 3740
OY 4435 ggaacgacagcagccgtatgtcaagtgatctgtgtgacaagagatctgatatgcaa 4494
|||||

Db 3741 GACACTGCCAGCACCATATGTCAGGTGTACTGTGTAGACAAACGAGATGTCATAGCCAA 3800
OY 4495 aaagaataaccagatggtggagagaagaacctgagccccgttaccagaagctctgtccct 4554
Db 3801 AAAGAAACCAAGGTGGCGAGAAAGACGTGGAGCCCTGTATACAGCAGCTGTATTCTT 3860
OY 4555 cgaagagagcccccgaggaggtgttacaagatcattgtctcgtggagatattgtcgtat 4614
Db 3861 CCAAGAGAGCCCCACAGGGGAGAGTGTACAGATCATTTGCTGGGAGATATATGCGCAT 3920
OY 4615 ggaacacaatcccttlatatggagtgccagatacctttagatgaactggaaactatccaa 4674
Db 3921 GGATCAACAATCTTTATGGAGTGCGCCAGATCTTATGATGAAGTGAAGTGAAGTATCCAA 3980
OY 4675 catgtgatgtgatgtgttcaaacctctccctccctccctccctccctccctccctcc 4734
Db 3981 CATGGTAAATGGATGGTGAACCTCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4040
OY 4735 acctctgacaagaagagcttcccaatcgltctctgagaagttctacccgacctcttactc 4794
Db 4041 ACCTTGACAAAGAGAGCTTCCCAATCGTCTGGAAGTCTACTGAGACTTCTTACTCTC 4100
OY 4795 tgcctcatagccaactataaactgtgtgcacaacaaccagcgatlacaaaaacagaagaa 4854
Db 4101 TCGTTCATAGCAACTATATAAAGCTGTGTCAACAACACAGCTGTACAAAAACAGAGAA 4160
OY 4855 aagcacaagtggaagccccctgtgaactgcattgttgaagtgtgtgtgtgtgtgtgtgt 4914
Db 4161 ACCACCCAGGTGG-GGACCCCTGCTCCACTGCATGCTGTGTTGTGTCTCCGGAGCCCA 4219
OY 4915 cgtctagagataccaagcgctcgtgtgtctcagagaag-----tcgtlacactgtgccc 4970
Db 4220 CCTCTTAGGATTAAGAGCCATCCTGTGCTCTAGAGGAAACCATCCGTGCACAGTGTGCC 4279
OY 4971 ctagcaaaagg 4980
Db 4280 CTAGCAAAAG 4289

RESULT 9
AF199327 LOCUS 5031 bp mRNA ROD 04-JUL-2000
DEFINITION Rattus norvegicus RIM2-4A (Rim2-4A) mRNA, complete cds.
ACCESSION AF199327
VERSION AF199327.1 GI:8925867
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 5031)
AUTHORS Wang,Y., Sugita,S. and Sudhof,T.C.
TITLE The RIM/NIM family of neuronal C2 domain proteins. INTERACTIONS
WITH rab3 AND A NEW CLASS OF src HOMOLOGY 3 DOMAIN PROTEINS
JOURNAL J. Biol. Chem. 275 (26), 20033-20044 (2000)
MEDLINE 20347919
REFERENCE 2 (bases 1 to 5031)
AUTHORS Wang,Y. and Sudhof,T.C.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1999) Center for Basic Neuroscience, The
University of Texas Southwestern Medical Center, 6000 Harry Hines
Blvd., Dallas, TX 75335-9111, USA

FEATURES
source location/Qualifiers
1..5031
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
1..5031
/gene="Rim2-4A"
52..4110
/gene="Rim2-4A"
/codon_start=1
/product="RIM2-4A"

[illegible]

1681 ttcactgagcctcgagagagctggtcattccacacctgagatatacaagctgtgtga 1740
1581 TTCACCTGAGCCTCAGAGAGAGAGAGTGCATCCACGCTGAGTATACAAAGCTGTATGA 1640
1741 tctgagcctgtaagagcgaagctgtgtgagtgtaagaagggagacatataaagggaaga 1800
1641 CGTGAAGATTGAAAGCCGAGACGTGATGAGAAAGAGACA----- 1681
1801 aactagtagagagaggtttgtcgattcttaaccacagctgtgagagaaagaag 1860
1682 ----- 1681
1861 gatgtactatgtgtgccaactcttggaaaggatttggaaatggtctgagcctcagattaa 1920
1682 -----TGAGTACAGACTGCTGGAGACATCCGCTTG 1712
1921 ggaactcgtggttagataactctgtagtagcaaaccttaacggagagcatagaaagtga 1980
1713 G-----CATGCACTGAGGCAATCCCAATGTC 1739
1981 taagcacccttgacctgagcagccatccaaagatgtgagatcgacctaatgtgtcgtattt 2040
1740 TTTGCACCCCTGTGACTGGACCAATCCAAAGATGAGATCGCTAATTGGTGTATT 1799
2041 attaaataagcgtttaaaagatgggaagtgtacctcgagactcaaggacatgtctggctt 2100
1800 ATTAAATAGCGCTTTAAAGATGGGAGTGTACCTCGATTCAGATTCAGAGCAATGCTGGCCTT 1859
2101 aaaggctttagaggaagaagatgaatcaggtcagctgtgcaattgtcattataccaagt 2160
1860 AAAGCTGTAGAGAGAAAGATGATGATCAGATCAGTGCACCTTGTGCAATTTATACAAAGT 1919
2161 aaaaaaggaagtttagctgtatctgtatagtagacatcttaagccaggtgtatgaagcttga 2220
1920 GAAAAAGGAAGTTTAGTGTGATCTAGTACATCTTAGACATCTTAGACAGTATGAGCTTGGA 1979
2221 atggaatgggaagctatgtgcaaggagccacatttggaaagtttaacaacttctcaga 2280
1980 ATGGATGGAAAGGCTATTGCAAGAGAGCCACATTTGAGAGATTTACAACTTTATTTCTGA 2039
2281 atccaaacctgagcaacaagtttgaagctgtgttccaaggccaatttggatatctctag 2340
2040 ATCCAAAGCCTGAAACCAACAGTGTGAGCTTGTGTTCAAGCCGATTTGAGATATGCTTAG 2099
2341 aatacctgtatgaacagatgacacaaactgtgaatccagttctagctcatattgaatcaaaa 2400
2100 AATACCTGTATGACACATGACACAACTGGAATCAGTTCTAGCTCATTTGAATCTCAAAA 2159
2401 aatgagcgtctctctctatactccgttaaccatagatccctgtgacatgtctgagagatgt 2460
2160 AATGGATGCTCCGCTATATTCGCTTACCTCTCCATGATAGTCTTGCGATGCTGAGGAGATGT 2219
2461 cccgagcttctatcttgcagacagcttccaataaactatgttltgacaaggttggtaacca 2520
2220 CCCACAGCTTCTATCTGACAGAGCTTCAATAAACAATGTTGTTGACAAAGTTGGTCAACA 2279
2521 gttgtagtatacaatttgggagcaaaagatctcccttcaagggaatgtggagagccaag 2580
2280 GTTAAATAGTTACAAATTTTGGAGCAAAAGATCTCCCTTCCAGGGAGATGGAGGCCAG 2339
2581 gaatccttctgttaagattactctctccagatagaagatgataaaaaataagagaagaac 2640
2340 GAATCTTATGTTAAATTTACTTCTTCCAGACAGAAATGATATAAACAAGAGAGAAC 2399
2400 AAAAACAAGCAAAACCTTGGAAACCAATGGAACACAGACTTTCATTTATCTCCCTGT 2459
2701 ccaacgaagaagaattcgtgtgaacgaatgctggaaatataccttggagataagagctagagt 2760
2460 TCACCAAGAAGATTCGGGAGCAATGCTGAAATCACACCTTTGGGATCAAGCTCGAGT 2519
2761 tcgagaagaagagagcaattctttagagagatttaattgaattgaaacagcttctgt 2820

2520 TCGAGAGAGAAAGATGATATTCCTTAGAGAGATTTTAATTGAATTTGAAACAGCTCTGT 2579
2821 agatgatgagccgcagctgtgtataagctgtgacaaccaaagtatgtctccatctgcaactccc 2880
2580 AGATGATGAACACACTGTGTACAAACTTCAGACCATGATGTCTCTCATATGCCACTCCC 2639
2881 tgcaccttcccatactcgtcccccgaagcagctccaatgtgagagagcccaagcagagct 2940
2640 TCACCTTCTCCATATATATGACCAAGAGACAGTGCATGAGAGAGAGCCCAACACGAGGTT 2699
2941 gcaaggctcgaagaagaatgaatgtacagatgaagtgctcgaactcgaactcgagatggagt 3000
2700 GCAAAAGCTCGAAGAAATAGATGACAGGCAATATCCGACTATGACTGGAGAGACGGCT 2759
3001 gggagtagtgcagattatcgacacaatgtgcccgaactctccaagctccaagctgttcggt 3060
2760 GGGAGTGTGTACATTTACCGACAGATGGCCGTGATCTTCAAAAGCTCCACATTTATCACT 2819
3061 gccagaagaatgaatgtatataaatcattgtccacatcagatgaggtctccatcgagtaga 3120
2820 GCCAGAAACAGTCATGATCATCAAAACCATTTGCTCACCATATGAGCTCTCTCATCGAGTAGA 2879
3121 tgtataaggaagacaagatcatgtgtcgactgtgccccctcctcctcaaggaaatgttga 3180
2880 TGTATTAGAAGAGATAGTATAGTATGCTGCGCTAGTGCCCTCCCTCAAGGAATGTGA 2939
3181 acaagggacacgagagacagctgtctactgtgccaattacaacaaatlaagccgaatgtatg 3240
2940 ACAGGGGCTTGGAGGAGACACGCTACTGCGCATTAACATTAATTAAGCCGATGAGACAG 2999
3241 aacccgttcaatgataagctacactctctcagatagagac----- 3280
3000 ACACGCTGATGACGACACATTTACTCTTCAGAGAGACACAGTCATTTCTCATCTAC 3059
3281 -----agggattgtgaagc 3294
3060 TCGCTCTGACACAGACAGACCATGAGCATCACCAAGGGATGAAAGGATTTGTANAGC 3119
3295 agcaatatgacagccatattcagaatccagatccaagaagaacagccctctcctagaagc 3354
3120 AGCAATATGACACGCAATATCACAGATTCAGATCAACAGACAAAGGCTCTCTTAGAGCG 3179
3355 gaccacacccgctccagatctcctgtgaagctcctgaatacaacctatgaagtgatgccc 3414
3180 GACCAACACCGCTCCAGATTCCTGTAAGTGCTGATPACAAACCTCATGAGTCAAGTCC 3239
3415 ttcaatgatgtgaagaatctgtccctctccaactgtcctatgtgaagtctaacctctg 3474
3240 TTCATTAATGACTGGAAGATCTGCCCTCTTCACCTGTGCTATGAGGCTCACCTCG 3299
3475 taacgggtctgtccagagaagaagcccatcaagttactccgggagacagagagggccgaca 3534
3300 CACTGGCTGTGTCACAGACAGTCCGTCAAGTACTCCCGTAAACAGAGAGAGGGCGCGCA 3359
3535 gcttccacagcttccaaccaagggaacatgtgagagaagtgctatgtatagaagagag 3504
3360 GCTTCTTCAGTCCCGACCAAGGAGACACGTGAAG----- 3395
3595 aaatgcacaatgaaacttaacaataatacaaacagtgtagccgatalcgaacccagactgga 3654
3396 ----- 3395
3655 gcaagatlaaccattggaaglatcgtccaggaatggatccacatagaagggcagatactgt 3714
3396 ----- 3395
3715 ttccactaaactcctcgagagatgatlgaatgtatctcggttltcaagagactagtag 3774
3396 ----- 3395
3775 tgccttcgttlttagcagcaacaagctacaatgtccgttccaatcagagcgccgagagga 3834

Db 3396 ----- 3395
Oy 3835 caggaaatcagtgctttataccatcaaatgcaaaacagacagatggcggtgctgggaa 3894
Db 3396 ----- 3395
Oy 3895 gaacttgacaaaagacacagcatcagtgagacatgctcacttgagagaatgacg 3954
Db 3396 ----- 3395
Oy 3955 cagccagctccagacactgacgtggcgccctgggtaccagtgagcaagaagcgagatcag 4014
Db 3396 ----- 3395
Oy 4015 cattggggccaatgtagctatgtgtgctctcaccgaaaagtcgcagtgctctca 4074
Db 3396 ----- 3395
Oy 4075 actcagccaacccgaaggaggtaaaaaagctacgagacactgttcagagaagcagga 4134
Db 3396 -----AGGAGAGAGTAAAAAGCTAACGACGACCTGTCCAGACAGACACGCA 3440
Oy 4135 gaccgggctgacgagtgatgagatgaggaactgtagacccgacagcgccagcggaatccac 4194
Db 3441 GACCGGGCTAGCAGTGGAGATGAGGAACCTGGATGACAAAGCAGCGCCGCGGATCCAC 3500
Oy 4195 agatggcagcatgacagcagctatagctcggaaggaatctgactctccctggggtccgct 4254
Db 3501 GGATGGCAGCATACACAGTTACAGCTCAGAGGAATCTGATCTTCTTGCGGCTCCGCT 3560
Oy 4255 ggcctctgacgacgagttcagtgatctctctgtagtgccctggccctgctcagctagtg 4314
Db 3561 GGCGCTTGACAGCGACGATGACGATTTCTTGATGCTGCGCCCTGCTCAGCTAGTGGG 3620
Oy 4315 aagcgaagccctggtcactccctgcaatgggtgacatcagtgagtggaatgatatgaa 4374
Db 3621 AGCCGACGACCTTGCGCAGCTCTCAATGGGTGACATTCAGGTGGGATGATGAATGA 3680
Oy 4375 gggagcagctgggaagtgaatactccggcgccgagcctgtgtgtaaacagagttccaa 4434
Db 3681 GGGACAGCTGGAGGTGAGATTCATCCGGGCTCGAGGCTTGATTAACACGAGTTCCAA 3740
Oy 4435 gacactgccaagcagctatgtccaagtgtagtctgttagacaacgagagttcagatagccaa 4494
Db 3741 GACACTGCCAGCAGCATATGTCAAGGTGTACCTGTGACACAGGAGTCTGATAGCCAA 3800
Oy 4495 aaagaaacccaaggtggcgagaaagacccttgagccctgtaccagagctctgtcctt 4554
Db 3801 AAAGAAACCAAGGTGGCGAAGAGAGCTGGAGCCCTGTATCAGCAGCTGTATCTTT 3860
Oy 4555 cgaaggagagccccaaggagggtgtttacagatcatgtctgaggagatatagtgctat 4614
Db 3861 CGAAGAGAGCCCGGAGGGGAGAGGTGTACAGATCATGTCTGGGAGATTAATGCGCGAT 3920
Oy 4615 ggaatcaaaatcctttatggagtgagccagatactctatagatgaactggaactatccaa 4674
Db 3921 GGATCACAAAATCTTTATGGAGTGGGCCAGATACCTTGATGATGAACGAGAACTATCCA 3980
Oy 4675 catgtgattgtgagttcaaacctctccctcctcctccctagtagatccaactcggc 4734
Db 3981 CATGGTAATTTGGATGGTTCAAACTCTTCCCTCTCTCCCTAGTAGATCAACCTTGCC 4040
Oy 4735 acctctgaaaagaagaggtcccaatgctctctgaaaagtctaccggaactcttactc 4794
Db 4041 ACCTCTGACAAAGAGGCTTCCCAATGCTCTCTGGAAAGTTCTACTGGACCTTCTTACTTC 4100
Oy 4795 tgcgttaatagcaactaataaactgtgtgccaacaacacagcagatcaaaaaacagaaga 4854
Db 4101 TCGTTCATATGCAACTATATAAACTGTGTGACACAAACACAGGTTACAAAACAGAAAGAA 4160
Oy 4855 aagcgaaggtggaagccctctgttaacactgcatgtctgtatgtgtgtctcagaagccca 4914
Db 4161 ACCACCCAGGTGG-GGAGCCCTGGTCCACATGCATGTGATGTTGTGTCTCCGAGGCCCA 4219

Oy 4915 cgtctagagataccaagcagctcgtgtctctcagaggaagc----tcgtacacatgtgccc 4970
Db 4220 CGGTCTAGGATTAAGAGCCACTCTGTGCTCTCAGAGGAAACCATCTGTCACATGTGGCC 4279
Oy 4971 ctagcaaaag 4980
Db 4280 CTAGCAAAAG 4289

RESULT 10

AF199332

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

BASE COUNT 1503 a 1166 c 1246 g 1116 t
ORIGIN

|||||
Db 1860 AAAGGTTGAGAGGAAATGACTGAATCAGCTGCACTTGTGCATTATTAACCAAGT 1919
Oy 2161 aaaaaaagaagttagcgtgatactgtagacacttagccagggatgaatgactgga 2220
Db 1920 GAAAAAAGAAAGTTAGCTGATACCTAGACACTTGAACCAAGTATGAAGCTTGGGA 1979
Oy 2221 atggaatgggaagctatctgaaggaagccaatctggaagttaacaactatctaga 2280
Db 1980 ATGGAAATGGAAGGCTATTTCAGAGAGCCACATTGAGGAAGTTACAACTATTATTAGA 2039
Oy 2281 atccaaactgaaccacaagtltgagctltgtlttcaaggccaattggaataltccag 2340
Db 2040 ATCCAAGCTGAAACCCAGAGTTGAGCTGTGTTTCAAGSCCGATTGAGATATGCTTAG 2099
Oy 2341 aatacctgtatgacaagcgtgacacaaactgtgaatccagttctaagctcatlttaactcaaa 2400
Db 2100 AATACCTGTATGACACATGACACACTGGAATCCAGTTTACGTCATTTGAATCTCAAA 2159
Oy 2401 aatggaacgctctctatactccgttaccctcaacatgagtcctggaatgctggaaggtgt 2460
Db 2160 AATGGAATGCTCCGCTATATTCGTTACCTTCCTCATGAGTCTGCGCATGCTGAGGATGT 2219
Oy 2461 cccgagttcttatactgagacagcttcaataaactatggtlttgacaaaggttggtacca 2520
Db 2220 CCCACAGTCTTATGTCGACAGCTTCAATTAATGTTGACAAAGTTGTCACCA 2279
Oy 2521 gttgtagtatacaatttgggagcaagatctccctccaggaagaatgggaagccaag 2580
Db 2280 GTTATATAGTTACAAATTTTGGGACAAAGGATTCCTTCCAGGGAATGGAGGCCAAG 2339
Oy 2581 gaatcccttltgtaagatttactctctccacagatagaaggttaaaataagaagaagac 2640
Db 2340 GAATCCTTATGTTAAATTTACTTCTTCCAGACAGAAAGATTAATAACAGAGAGAAC 2399
Oy 2641 aaaaacagtcaagaanaactltggaacccaatltgaaacagacttcaattatctcgt 2700
Db 2400 AAAAACAGTCAAGAAATCTTGGAAACCAATGGAACCAAGACTTTCATTATCTCGT 2459
Oy 2701 ccaacgaagaagaatctcgtagaacgaatgtctgaaatcaaccttggagtaagtaagt 2760
Db 2460 TCACCGAAAGAAATTCGGGAAAGAAATGCTCGAAATCACCTTGGGATCAAGTTCGAGT 2519
Oy 2761 tcgagaagaagaagaagcaattcttagagaagatttaattgaatttgaagaagcttgct 2820
Db 2520 TCGAGAGGAAGAAAGTATCTTAGSAGAGATTTTATTTGAATTTGAACAGCTCTGTT 2579
Oy 2821 agatgtagccgagcttggatataagctgacagaccatgactgtcccatctgcaactccc 2880
Db 2580 AGATGATGAACCACTGGACAAACTTGACCAATGATGTCCTCATTTGCCACTGCC 2639
Oy 2881 tcgaccttcccatatctgcccggagcgagctccatgtagagaagcccaagcgaggt 2940
Db 2640 TCACCTTCTTCATATATTCACGAGACAGTGCATGGAGAGGCCAACACCCAGGTT 2699
Oy 2941 gcaaaagtgcgaagaataagtgacagtggaagtgtctgactacagcgcgaagatgagct 3000
Db 2700 GCAAAAGTGAAGAGATATATGACACAGCAATATCCGACTATGACTGCGAGGCGGCT 2759
Oy 3001 gggagtagtgcagattatctgacacaatggcgagatcttcaaaagctccaagctgtgcgt 3060
Db 2760 GGGAGTCTGTGAGATTACGACAGCATGCGGTGATCTTCAAAAGTCCACATATATAGT 2819
Oy 3061 gccagaacaagatcgtcatcaaatcatgtctcaacatcaaggtctccatcgataga 3120
Db 2820 GCCAAACAAAGTATGCTCATCAAAACCATGTCTACCATCAAGGCTCTCTCATCAAGTGA 2879
Oy 3121 tgttaagaagaagaagaatcgtgctgagccttagtgcctccctccctcaaaagaaatgaga 3180
Db 2880 TGTATAGGAAGACTAGGTGATGCTGCTTGTGCTTCTCTCTCTCAAGGAATGTGGA 2939
Oy 3181 acaggggcaacgaggaacagtgctactgagccattacaacaatltgacgaatgtag 3240
|||||

Db 2940 ACAGGGGCTTCGAGGACACGTGCTACTGCGCATTAACAATACATTAGCCGATGACAG 2999
Oy 3241 aaccgtgctcatgagatgagaccactactcttcaagatagagac----- 3280
Db 3000 ACACCGTGTATGACGACGACCATTTACTTCTTACAGAGAGACAGTATTTTCTACTTACC 3059
Oy 3281 -----aggatltgaaag 3294
Db 3060 TCGCTCTGCACACAGGCACACCGCTGAGCATCACACAGGAGTGAAGGATGTGTAGAC 3119
Oy 3295 agcagatagacagcaatcacaagatccagatcaacagaacaaagcctctcctagagc 3354
Db 3120 AGCAGATAGACAGCCATATACAGATTCAGATCAACAGAAAGAGCGCTCTCTTACAGAGC 3179
Oy 3355 gacacacacccgctccagatccctctgaagctccgtgtatacaaaactatgaggtgtgagcc 3414
Db 3180 GACCAACACCCCTCCAGATCCCTGAACCTGTGATACAAACCTCAATGAGGTGAGATGC 3239
Oy 3415 ttcattaatgactggaagatctgcccctccctcaactgtcttalcgaaggtcaccctcg 3474
Db 3240 TTCAATTAATGACTGGAAGATCTGCCCTCCTTCACTGCTTATCGAGTCTCACCTCG 3299
Oy 3475 tacgggtctgtccagacaagccatcaagtaactccgggaaacaggaagggcgaca 3534
Db 3300 CACTGGCTCTGTCCAGACAGCTCCGTCAACTACTCCCGTAAACAGAGAGAGGCGCGCA 3359
Oy 3535 gcttcacagcttccacaaaggaacatltggaagaagtgtcatgtatatagaagagag 3594
Db 3360 GCTTCTCAGCTCCACCAAGGGAACTCGAAG----- 3395
Oy 3595 aaatcgccaatlgaaacttaacaatatcaaaacaggtgacggatcagaacccagatcga 3654
Db 3396 ----- 3395
Oy 3655 gcaagatcacatcgaaglatcgctcgaagaatggaatccacatagagggcaatcagt 3714
Db 3396 ----- 3395
Oy 3715 ttcacataaactctcgagacgtgataagtgatgatactcgcggttccaagactagtag 3774
Db 3396 ----- 3395
Oy 3775 tgcctcgttccagcagaagaactatcgtccgtccaaatcagagcggccgagaggaa 3834
Db 3396 ----- 3395
Oy 3835 caggaanaatcagtgcttcaacatccaaatgcaaaacagacagatgggcygtgcgggaa 3894
Db 3396 ----- 3395
Oy 3895 gaacttgacaaaagacacacgactcagtggaagacatgtgtcactgtgaagaagaatgag 3954
Db 3396 ----- 3395
Oy 3955 cagccagtcgcgaactgcagtgggcgccctgggttacacagtgycgaagaagcgcgatcag 4014
Db 3396 ----- 3395
Oy 4015 catlgyggccaaaatgtagctaltgtlgtgtctctcagcggaaaagtcgagtgctctca 4074
Db 3396 ----- 3395
Oy 4075 actcagccaaacccgaagggaggttaaaaagctacaggaagacgtgttaagaagaacagga 4134
Db 3396 -----AGGAGGAGGTAAAAAGCTAACGACGACACTGTCCAGAGAAACACAGGA 3440
Oy 4135 gacgggctagcagtgtagagatgaggaactgtgaacccgcgaagccgagccgggaatccac 4194
Db 3441 GACCGGCTTAGCAGTGAATGAGATGGAATCTGATGACAAAGACAGGCCGCGGGAATGCAC 3500
Oy 4195 agatgacagcatgaacagctatagctcgaagaagaaatcgtactcttccctgggtcgct 4254
Db 3501 GGATGGCACATGAAACAGTTACAGCTCAGAAAGAAATCTGATCTTCTGCGGCTCGCCT 3560
|||||

| | | | |
|------------------|--|--|-----------------|
| OY | 4255 | ggcctctgaaagccagttcagtgatcttcctggaagcgcttgcctcaagttaagg | 4314 |
| Ddb | 3561 | GGCCTCTGAAGCCAGTTCACCGATTTCTTGATTGTCGGGCCCTGTCAAGTAGTGGG | 3620 |
| OY | 4315 | accgcacacctgtgctactctctgcaatgggtgacaattcagctggaatatgatataaaa | 4374 |
| Ddb | 3621 | ACCGCACACCTTGGCGACTCTCTGCAATGGGTGACATPTCAAGTGGGATGATGATAAAGA | 3680 |
| OY | 4375 | gggacagctggaagtagaataatcatcvcggcgccgycgcgccttgygtataaaccaagttccaa | 4434 |
| Ddb | 3681 | GGGACACTGTGAGGTAGATCATCCGGGCTCGAGCCCTTGTAAGTAACCAAGTTTCAA | 3740 |
| OY | 4435 | gaacactccaagcacctgtaattgcaaggtgtatcttgttagacaacggaagcttgcataagccaa | 4494 |
| Ddb | 3741 | GACACTGCACACCATATATGTCAAGSTGTACTGTTAAGAACAGGAGTGTGCATAGCAA | 3800 |
| OY | 4495 | aaagaanaaccaagtglygcgaagaagaacctggaagcccctgtaccagaagctcttgtcctt | 4554 |
| Ddb | 3801 | AAAAGAAAAACAAGGTGGCGAGTAAGACGCTGGCCCCCTGTATCAAGCATGATATCTT | 3860 |
| OY | 4555 | cgaagaagaagcccccaagggaagggtgttacagatcatctgtctgggagaatlatagtgctat | 4614 |
| Ddb | 3861 | CGAAGAGAGCCCCCAGGGGAAGSTGTTACAGATCATTTGTTGGGAGATTAATGSCCGAT | 3920 |
| OY | 4615 | ggatcacaatccctttaatggaagtggcccaagatacctctatagatgaactggaactatccaa | 4674 |
| Ddb | 3921 | GGATCAACAATCCTTATAGGGAGTGGCCCCAGATACATCTTATGATCAACTGTGAATATCCA | 3980 |
| OY | 4675 | catgttgattgtaigtgttaaactctccctctcctctccctaagtagatccaacctgcgc | 4734 |
| Ddb | 3981 | CATGTTAATTTGGATGGTTAAACTCTTCCCTCCTCCTCCTAATAGATCAACCTTGAC | 4040 |
| OY | 4735 | aaccttgacaagaagagcttcccaatcgctctctggaagttctacccgaactcttctact | 4794 |
| Ddb | 4041 | ACCTCTGACAAAGAGAGCTTCCCAATCGTCTCTGGAAAGTTCTACTGGACCTTTACTC | 4100 |
| OY | 4795 | tgcgtcataggaactataaactgctgtctcaacaacaacagcgatatacaaaaaaccaagaag | 4854 |
| Ddb | 4101 | TGCTTCATAGCAACTATAAAGTGTGTCTCACAAACACAGSGTTATCAAAAAACCAAGANA | 4160 |
| OY | 4855 | aacgcacaggttgaaagccccctgyltaaacatgcatgtcttatgttgtgtctacagaagccca | 4914 |
| Ddb | 4161 | ACCACCCAGAGTGG-GGACCTCGTGTCCCACATGCAATGCTGATGTTGTCTCGGAGCCCA | 4219 |
| OY | 4915 | cgtctaggagataccaagaagcttcctgtgtctccagaagaag-----tcgtataacattgtgcc | 4970 |
| Ddb | 4220 | CGTCTTAGGGTAAAGACCATCTCTGTGCTCTACAGAGAAAGCANTCGTGCACAGATGTGCC | 4279 |
| OY | 4971 | ctagcaaaag 4980 | |
| Ddb | 4280 | CTAGCAAAAG 4289 | |
| RESULT 11 | | | |
| LOCUS | AF199325 | 5079 bp mRNA | ROD 04-JUL-2000 |
| DEFINITION | Rattus norvegicus RIM2-3A (Rim2-3A) mRNA, complete cds. | | |
| ACCESSION | AF199325 | | |
| VERSION | AF199325.1 GI:8925863 | | |
| KEYWORDS | Norway rat. Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; | | |
| SOURCE ORGANISM | Rattus | | |
| REFERENCE | 1 (bases 1 to 5079) | | |
| AUTHORS | Wang,Y., Suiita,S. and Sudhof,T.C. | | |
| TITLE | The RIM/RIM family of neuronal C2 domain proteins. INTERACTIONS WITH rab3 AND A NEW CLASS OF SRC HOMOLOG 3 DOMAIN PROTEINS | | |
| JOURNAL | J. Biol. Chem. 275 (26), 20033-20044 (2000) | | |
| MEDLINE | 2034/919 | | |
| REFERENCE | 2 (bases 1 to 5079) | | |

| | | | |
|----|------|--|------|
| OY | 361 | gcaagtcgaagaagatcgggagaggaatcgcagcagcgcagagcagaagggtgaatgcgcc | 420 |
| Db | 261 | GCMAATGCAAGAAAATGGGAGAGAAATTCACAAACAGCAGCAAGACAAAGGCGCACGCC | 320 |
| OY | 421 | gaccctgagcatctcccaagaacaaattcttcagatgtagtgcgccataatgtgttccta | 480 |
| Db | 321 | GACCTCGGGCATCTCTCCACACACAAAATTTGCAGATGCTGCGCCATTACTGTTTCATA | 380 |
| OY | 481 | ttgccaacccaagttctctgtccctcgaatgtgagagtcgaagttccttcaagtcacaaga | 540 |
| Db | 381 | TTGCCAACAACAAAGTTCTGTGCTGTGTGTGAGAGTGCAGTGTATTCAGCTCAACAAGT | 440 |
| OY | 541 | tatgtgggtgtgtaatttgttcgcgaacaacaagaatctcactaataatcagagcatg | 600 |
| Db | 441 | TATGTGGGTGTGTAATTTGTGCGCGAAMACACAGAAATCTCTCAATGAGCGGCGGTG | 500 |
| OY | 601 | gtttataatagtggtctcaaacacttcagcaaaccttgctcaaaagttctctgagagct | 660 |
| Db | 501 | GTTTACATATAGGTGGTCTTAATACCCGACGCGCTGTATCAAAAGCTTTGAGGGCT | 560 |
| OY | 661 | tcgaatctagaggaagcccttcagagagaagaacaaactacacagacagagcccaagttcca | 720 |
| Db | 561 | TCGCAGTGAGGAAGCCCTCAGAGAGAAAGGCAAAACGTGCATGACAGACGAGTTTCCA | 620 |
| OY | 721 | aggagcccccaggttgacttaccgtactctcgaagttgagaaagccgagctcctatggctcac | 780 |
| Db | 621 | GGGACCCCCCGGTGACTCATGTACGTCTCAGTTGAGAGAGTGTGAGCTCATGGGCTCAC | 680 |
| OY | 781 | aagacaagatctatttaaaatggtatccagagaggaagcaccbaattggcagttacatgcc | 840 |
| Db | 681 | AAAGACAGGATTTCTAATTAAGAAATGGCTCAGAGAAAGCAACCAATGGCCAGTCACATGCC | 740 |
| OY | 841 | ttcagacagaaaacggaagtcacatcgaatgctccagaggaatcaaaatcgaagatatacgacaag | 900 |
| Db | 741 | TTCCACAGCAAAAAGAAAGTCCTGCATGTCTCCAGAGATCAAAATCGAAGATACACCAAG | 800 |
| OY | 901 | tgaagaaagagsggactactcaagatgttccctcgaatggtacaaatgccaagatctcc | 960 |
| Db | 801 | TGAAGAAAGAGAGCAATATTCCACGATATGTTCTCTCAGATAGCACAATGCCCTGATCTCC | 860 |
| OY | 961 | ttcggatattgtctgaagcagatctcaggtgtagcctcaatttatgaanaacctgtca | 1020 |
| Db | 861 | ATCCAGATTTATGCTGATAGACGATCTCACGCTGAACCTCAATTTATGAAACCTGATVCA | 920 |
| OY | 1021 | tttaaatcacaaggtatccaagaagagagccatagacatcccaagatataitgga | 1080 |
| Db | 921 | TTTAATATTATAGGATTTCTAACGAGAGAGCCCTATACATTTCCAAACGATATTTGTAGA | 980 |
| OY | 1081 | tgaatgaagatgtggagagcagagatgaaatatagaagaacaaaggagagaggaatacca | 1140 |
| Db | 981 | CGACGAAAGATGTGGAGAGCAGAGATTAATATGAAACAAAGGAGAGAGGAGTGATVCA | 1040 |
| OY | 1141 | ggcagcgtacagaatggtatccaaatcttgccccgtatcccgtaagcccaacctcagca | 1200 |
| Db | 1041 | GGCAGCGTACAGAAAGTGAATCCAAATTTGGCCGGTATCCGGTAAAGCCACACCCATATGA | 1100 |
| OY | 1201 | agaacaatctgcattccacgcgtbaggtgtcccaaggagacacatgaagagagagcaagtga | 1260 |
| Db | 1101 | AGAGCAAAATCGCGATCCACGCTGAATGTGCCGGGACACGACAGAGAGAAAGCACATGA | 1160 |
| OY | 1261 | tgttctcttgccaacgcgtgaactagaaatctcagaattctctgtctaaagatgtagatg | 1320 |
| Db | 1161 | TGTTTCTTTGGCAAAAGTGTGAATTAAGATTTCCAGAGATTTCTGTGTAGAGATGATAG | 1220 |
| OY | 1321 | accatcaggagcaagaatctgtatctgaacgttagagctgcaatggaanaacaaacagatcgta | 1380 |
| Db | 1221 | ACCATTCAAAGCAAAAGTCTGTATCTTAACGTAAAGCTCAATGCAAAACACAGCATGTGA | 1280 |
| OY | 1381 | ttcaatggaagaacatccgagaggtccaggaacaaagttctatccaaanaagacctcaaa | 1440 |
| Db | 1281 | TTCAATGCAAAAGACTCGAGAAAGCTAGGAGCAAAAGTTCTTATTCACAAAGGACACCAAA | 1340 |
| OY | 1441 | tcaatgctctccaccccttcggcgagacctataccgcttgatagaccagacatlgagcg | 1500 |

[illegible]

```
Db 2280 TGTTCCTAGGGTTTCAGATATAAACTATGTTTGACAAAGTTGGTCAACCACTTAATAGTTAC 2339
Oy 2533 aatttgggggcaaaagatctccctcccaaggagatggagggcaagaatccctatgt 2592
Db 2340 AATTTTGGGAGCAAGATCTCTCTCCAGGGAAGATGGGAGCCAAAGAACTCTTATGT 2399
Oy 2593 taagatttactctccctccaga tagaagtgtataaanaa taagagaagaacaanaaasagtc aa 2652
Db 2400 TAAATTTACTTCTCTCCAGACAGAGTGAATAAAAACAAGANMAACAAAAACAGTCAA 2459
Oy 2653 gaaaacttggaaecccaatgtagaacccagactttaattatctctgtccccaagagaga 2712
Db 2460 GAAAACTTTGGAMCCCAATATGGAACCAAGACTTTCATTATTTCTCTGTTCCACGANAAGA 2519
Oy 2713 attccggagaagaaatgctggaataatccctltggatcaagcttagagttcgagaagaaga 2772
Db 2520 ATTCCGGGAACGAATGCTCGAATTCACCCCTTGGATCAAGCTCGATCGAGAGAGAGA 2579
Oy 2773 gaacgaattcttaaggagaatlttaattgaattggaacaagcttgtctagatgtagagcc 2832
Db 2580 AAGTAATTCCTAGAGAGATTTTAAATGAATTGGAACAGCTCTGTTAGATGATAACC 2639
Oy 2833 gcaatgtatagcttgagaaacccaatgatatgtctcccatgycacatccctccgcccctccc 2892
Db 2640 ACACTGTACAAACTTCAAGACCCATGATGTCTCTCATTTGCCACTCCCTCACCCCTCTCC 2699
Oy 2893 atactcggccggaggagcagctccatctgagagagaccgaacgagcagctgcaaaagtctgaa 2952
Db 2700 ATATATGCCAGGAGACAGAGCTCATGAGAGAGACCACAGCAGAGTGTCAAAAGGTCGAA 2759
Oy 2953 gagaataagtacagtagaagtgtctgactacagctgagagagatgggttggaatagtgtc 3012
Db 2760 GAGATAATAGTGACACGGAAGTATCCGACTATATACGTGAGAGAGGCGGTGGAACTCGTGC 2819
Oy 3013 agattatgacaacaatgagccgagctcttcaaaagctccacagttgtgcgtgagcaagaagt 3072
Db 2820 AGATTATCCGACACGATGGCGGTGATCTTCAAAAGCTCCACATTATTCAGTGCAGAACAGT 2879
Oy 3073 catgtcatcaaatcatctgtctcaacatcaaggtctcctcatcgatgtagatgtttataagaag 3132
Db 2880 CATGTCAATCAAAACCAATGCTCACACATCAGGCTCTCCTCATCGAGTAAATGTTATAGAG 2939
Oy 3133 gacaaggtcatgtgtgcctagatgagctcctcctcctcaaaagaaatgtggaagggcagcg 3192
Db 2940 GACTAGGTATGTATGTGCTTATGTATGATGATGATGATGATGATGATGATGATGATGATG 2999
Oy 3193 agggacacgtgtctactgagccatcacacaacaalltagccgaatgtagatacagcgtgtcat 3252
Db 3000 AGGGAACAGTGTACTGGCCATTCATCAATTAAGCCGGATGGACAGACACCGTGTCTAT 3059
Oy 3253 ggaatgacacactactctttagatagagac----- 3280
Db 3060 GGACGACCAATTAATCTCTGAGAGAGACAGTCAATTTTCTCACTACTGCTCTCGACA 3119
Oy 3281 -----agggaattgtgaagcaagcagataagaca 3306
Db 3120 CAGGCGACCAAGTAGACATCCACACAGAGGATGGAAGGATTTGTGAACAGACAGATAGACA 3179
Oy 3307 gscataatcaagatccagatcaacagaacaagggcctctcccttagagcgagaccacaacccg 3366
Db 3180 GCCATATCAACAGATCCAGATCAACAGMAACAGGCTCTCTTAAGGCGGACCAACCCCG 3239
Oy 3367 ctccagatctctggaagctctgatacaaacctcctagagtcgagtcgcttcaatgac 3426
Db 3240 CTTCACATCTCTGAACTGTCTGATCAAACTCATATGAGTCGATGCTTCAATTATATGAC 3299
Oy 3427 tggaaagatctgcccctccctccacactgcttatacgaggtctcaccctcgtaagggctgtgt 3486
Db 3300 TGGAAATATCTGCCCTCTCCCTTACCTGCTTATCGAGGCTCAACCTCGCACTGCGCTGTGT 3359
Oy 3487 ccagacaagcccatcaagtactcccggaacaggaacggaagggcccgacagcttccacaact 3546
Db 3360 CCAGACCAGTCCGTCAAGTACTCCGTAACGAGGAGGAGGCGGCGAGCTTCTCTCACT 3419
```

```
Oy 3547 tccaccaaaaggaacattggagagaagtgtctatgtatatagaggagagaatctgccaaat 3606
Db 3420 CCCACCAAAAGGAACACTGGAAAG----- 3443
Oy 3607 gaacttaacaatatcaacaacagtagccggaatcagaccaccaacttgagcaagattacca 3666
Db 3444 ----- 3443
Oy 3667 ttcgaagatcgtctcaagatgtagtaccacatagaggggcaataactgttccactaaatc 3726
Db 3444 ----- 3443
Oy 3727 ctggagacagtgatgtaagtgtatctgtcggttccaagactagtagtgcctcgtt 3786
Db 3444 ----- 3443
Oy 3787 cagcagacaagctacatgtccgtccaaatcagagcgccgagaggaagaacaagaaatcag 3846
Db 3444 ----- 3443
Oy 3847 tgtcttaacatccaaaatgcaaaaacagacagatggtgtgtcgtgggaagaaacttgacca 3906
Db 3444 ----- 3443
Oy 3907 aagcacacacatcag tggagacatgtgtcacaatggaagaatlgacgagcagccagctcga 3966
Db 3444 ----- 3443
Oy 3967 cactgcagtggtcgccctggtgtaccagtggtgcaagaagcggtcagatctagatctggggcga 4026
Db 3444 ----- 3443
Oy 4027 aatgttagctatgttgtgtctcctcaagaaagtgcagtgctcctcaactcagccaac 4086
Db 3444 ----- 3443
Oy 4087 cgaagagaggtaaaaaagctacgagacgtgtcagagaagcagaggaacgggtgtagc 4146
Db 3444 ---ACGAGAGAGTAAAAAGCTACGACGACTGTCCAGAGAACGACGAGCCGCGGTAGC 3500
Oy 4147 agtgaagatggaagactgtatgatacccgccagccagccagcggaatccacaagatgacagat 4206
Db 3501 AGTGAAGATGAGAACTGGATGATCAAGACAGGCCACCGGGAATCCAGATGGACAGAT 3560
Oy 4207 gaacagctatagctcggaaagaaatctgaatctccctggtgtccgctgtgcctgtcag 4266
Db 3561 GAACAGTTACAGCTCAGAAAGAAATCTGATCTTCTTGCGGGCTCGCCTGTGACAG 3620
Oy 4267 ccagttcagtgatctcctgtgtagtggtcgtggtccctgtctcaggtagtggaagccagaccct 4326
Db 3621 CCAGTTCACCGATTTCCTTGATGCTGTGGCCCTGCTCACTAGTGGAGCCAGACCTT 3680
Oy 4327 ggtctactcctcgaatggtgtgacatcaggtggtgaatgatatgataaaaaaggaacgttga 4386
Db 3681 GCGACTCTGTCATATGGGTGACATTTCAGTGGGGATGATGATTAAGAAAGGAGACGTGGA 3740
Oy 4387 ggtagaatcatcccggtcggtcggtctgtgtgtaaaacaggttccaagacactgacagc 4446
Db 3741 GGTAGAGATCATTCGGGCTCGAGGCTTGTAGTAAACCAGGTCCAAAGACACTGCCAGC 3800
Oy 4447 accgtatgtcaaggtgtatctgtttagacaacggagttgtgatalagccaaaaaagaacaaca 4506
Db 3801 ACCATATGTCAAGGTGTACCTGTTAGCAACGGAGTGTGATAGCCCAAAAAAGAAACCAA 3860
Oy 4507 ggtgtgagaagaagaccttgagccctgtacacagcagctctgtccttcgaggaagccc 4566
Db 3861 GGTGGGAGAAACACGTGGAGGCCCTGTATTCAGACAGCTGTTATCTTCCGAAGAGAGCCC 3920
Oy 4567 ccagggaggggtgtllacagatcatgtctgtgggagatlaagtcgtgtatgatacaaatc 4626
Db 3921 CCAGGGGAAGGTATTACAGATCAATTTGTCTGGGAGATTATGCGCATGATCAACAAATTC 3980
```


Db 681 AAGACGATTTCTATTAGAATGGCTCAGGAATGAAGCACCAGATCCGAGTGCATGCC 740
Oy 841 ttcagacgaaaaaagagctccatcagctgtccagggatcaaatcgaagatagcagcaag 900
Db 741 TTCAGACAAAAAAGAGTCCTGTCAGTGTCCAGGGATCAAAATGAAAGATACGACCAAG 800
Oy 901 tgaagaagagagactactacagatagtccttcagatggtlaacaatgccagatctcc 960
Db 801 TGAAGAAAGAGAGANTATTCCAGCTATGTTCCCTTCAGATAGCAAAATGCCATGATCC 860
Oy 961 ttcgagatagctgtagagacgctcagcgtaggcctcaatttatagaagacctggtca 1020
Db 861 ATCAGATTATGCTGATAGAGCATCTCAACGTGAACCTCAATTTATGAGAACTGATCA 920
Oy 1021 ttttaaatcagagatctcctacagagagagccatagacatccaaagagatattgtga 1080
Db 921 TTTAAATTATAGGATTTCTAACAGAGAGGCCATAGACATTCCAAAGAGTATATGTAGA 980
Oy 1081 tga tgaagatgtygagagcagaga tgaalatgaagaacaaagagagagaggaatacca 1140
Db 981 CGACGAAGATGTGGAGAGCAGAGATGATATGAAGACAAAGGAGAGAGAGATACCA 1040
Oy 1141 ggcacgctacagagatgatccaaatctgcccgtatcccgtaaagccaacacctaca 1200
Db 1041 GGCACGCTACAGAGATGATCCAAATTTGGCCGGTATCCGGTAAAGCCACACCTATGA 1100
Oy 1201 agaacaaatcgcaatccacgcttgagtggtccagggcagcaca tgsagagaagcacagtga 1260
Db 1101 AGAGCAATATGGAGTCCACGCTGAAGTGTCCGGGACAGACACGAGAGACACAGTGA 1160
Oy 1261 tgttctcttgcaaacgctgaactagaaagattccagagattctcgtcgaagatgtaga 1320
Db 1161 TGTTTCTTTGGCAATGCTGAAGTGAAGATTCCAGGATTTCTGCTTAAGATGATGAG 1220
Oy 1321 accatcagagcaaatgatctgtatctgaacgttagagctgcgaatgysaaaacaaagatcgta 1380
Db 1221 ACCATCAAGCAAAAGATGTTGATCTGAAGCTGAGCTGCAATGGAAACACAGGATCGTA 1280
Oy 1381 ttcaatgysaaagaaactcgagagagctcagggagacaagttcttatccaaagagacctcaaa 1440
Db 1281 TTCAATGAAAGAACTCCAGAGAGCTCAGGGACAAAGTTCTTATCCAAAGGACCAACAA 1340
Oy 1441 tcaatgtctccaccctccgagagccctataccgcttgatagacgaaactgaagcg 1500
Db 1341 TCATATGCTCTCTACCCCTCGCAGAGACCCCTATACGCTCGATAGACCAAGAGCTGAGCG 1400
Oy 1501 cgtcgaactccctacgysaaacagcaccaactlagatccagctctgctgtgagysaaacgaa 1560
Db 1401 TGCCGACTCCCTACGSAACACACACTTAGATCCAGCTCTGCTGTAAGGAAAGAA 1460
Oy 1561 gcgagaaaaaaatgysaaaccatgtttaagagatgattcttgagttcagacagtcagagtc 1620
Db 1461 GCGAAGAAAAATGSAACATGTTGAGGAATGATCTTTGAGTTCAAGACCAAGTCTGAGTC 1520
Oy 1621 agtggagccgcccccaacaaagccctcataaatccaaagaaagagatanaa tgcgccaagt 1680
Db 1521 ACTGAGCGCCGCCCCACCAAGGCTCATTAATCCAAAGAAAGAGATAAATGCGCCAGGT 1580
Oy 1681 ttcactgagcagctcgagagagagagctgtgcattccacaacctgaglatatacagctgtgtga 1740
Db 1581 TTCACCTGAGCACCTCTGAGAGAGAGATTGGCATCTCACGCTGAGATATCAAGCTGTGATGA 1640
Oy 1741 tgtgagctcgysaaagcgagagtg tlgagtgysaaagggagacagttcaaaagggysaaagaaa 1800
Db 1641 CGTGGAGATTGAAAGCGAGAGCTAGTGTGAAAGAGGACAT----- 1682
Oy 1801 aactagtgagcagagagtttltgcgattctlaaacaagagctctgagagacaaaagaaag 1860
Db 1683 ----- 1682
Oy 1861 gatgtaactatgtgtgccactcttctgysaaagagatttggaatggtctgagctcagattaa 1920
Db 1683 -----GAGTACAGCTGTGTTGAGAGCATGCGCTTTG 1712

Oy 1921 ggaactctgggttagatactctgttagtcacaaaccttaacagagagatagccaatgta 1980
Db 1713 G-----CATAGCATGAGGACATCCCAATGTC 1739
Oy 1981 taagcacctgtgactctgcaagcatccacaaga tggagatcgacttaatltgtcgatltt 2040
Db 1740 TTTGCACCTGTGACCTGTGAGGACATCCAAAGATGGAGATCGCCTAATTGGTGTATTTT 1799
Oy 2041 attaaatagcglttaaaaga tgggagtg ttaactcgagactccggaaacatgctgggctt 2100
Db 1800 ATTAATAAGCCGTTTAAAGATGGAGTGTACCTCGAGATTCAAGACCAANTGCTGGCTT 1859
Oy 2101 aaagtgtytagggysaaaga tgaactgaatcgagtcgagcttltgtcatltaaccaagt 2160
Db 1860 AAGGTTGTAGGAGAGAAAGTATGACTGATCAGTGCAGCTTGTGCATTTATTACCAAGT 1919
Oy 2161 aaaaaaagaagtttagctgatactgtagacatcttagaccaggtgataagctltgga 2220
Db 1920 GAAAAAAGAA GTTACCTGATCTGTAGGCATCTTAGACCAGGTGATGAAAGTCTTGA 1979
Oy 2221 atggaatggagagctatctgcaagagccaatlttgaggaagtttaacaactatctaga 2280
Db 1980 ATGGAATGAAAGCTATTTGCAAGGACACATTTGAGAACTTTNCAACATTATTCTAGA 2039
Oy 2281 atccaaacctgaaccacaaglttgagctgtgttltcaagccaatttgagabaltcctag 2340
Db 2040 ATCCAAACCTGAAACACACAGTTGAGCTTGTTTCAAGGCCGATTTGAGATATGCTTAG 2099
Oy 2341 aatacctgataagcagcatgcacacatggaatccaagtcttgatatttgatctcaaa 2400
Db 2100 AATACCTGATAGCACCTCATGACAACTGGAATCCAGTCTAGCTCATTTGAAATCTCAAAA 2159
Oy 2401 aatgagcgttccctctatctatccgttaccctcaacca tgaactcgtgagcagtgagagatgt 2460
Db 2160 AATGATGCTGCCGTATATGCTTACTCTCCATGATGCTCTGCAATGCTGAGGATGT 2219
Oy 2461 cccgagttcttlatctgagcagcttccaataaactatggtltgcaaggtgtgtcaca 2520
Db 2220 CCCACAGTTCTTACTGTGACAGCTTTCAATAAAMACTATGTTTGAACAAGTTGGTCCACA 2279
Oy 2521 gttgagttgaactlttgggagcaagagatctccctccagysagatlgysagccaa 2580
Db 2280 GTTATATGATTCAATTTTGGGACCAAGAGATCTCCCTTCAGAGAAAGATGGAGGCGCAAG 2339
Oy 2581 gaatccatgtttaagatttacttccctccagatagaagtgataaaataagagaaagac 2640
Db 2340 GAATCCTTATGTTAAATTTACTTCTCCAGACAGAAAGTGAATAAAMCAAGAGAAAGAC 2399
Oy 2641 aaaaacagttcaagaaactltggaacccaatgysaacagacttcatattatctcgt 2700
Db 2400 AAAAACAGTCAAGAAACCTTTGAAACCAAAATGGAACCAAGACTTTCATTTATTCCTGT 2459
Oy 2701 ccaacgagagagatctcg tgaacgaaatgctgysaaalttaacctltggatcaagctagat 2760
Db 2460 TCACCGAAGAGATTTCCGGAGACGAATGCTCGAAATCACCCCTTTGGATCAACACTGAGT 2519
Oy 2761 tcgagaagaagagogaattctctagggagagatttaattgaatttgysaaacagcttgc 2820
Db 2520 TCGAGAGGAAGAAAGTGAATTTCTTAGAGAGATTTTAATTTGAATTTGAAACACACTCTGT 2579
Oy 2821 agatgatgagcgagctgtgataagctgagaccaca tgaatgtctccatltgcactcc 2880
Db 2580 AGATGATGAACCAACTGTTACAACTTCAGACCCATGATGATGTCCTGCAATTTGCCACATCC 2639
Oy 2881 tgcacctcccatatctlgccccgagagcagctccatlgysagagagcccaacgagagct 2940
Db 2640 TCACCTTTTCCATATATGCGACGAGACAGCTGATGAGAGAGCCCAACAGCAGAGTT 2699
Oy 2941 gcaaaagttcgaagaagataagtgacagtgaaagtctgataagactgagagagagagct 3000
Db 2700 GCAAAAGGTGGAAGAGATTAAGTGACAGCGAAAGTATCCGACTATGACTGCGAGGACGCGCT 2759

| | | | |
|----|------|--|------|
| QY | 3001 | gggagtagtgtgaattatcgcacaatgycgcgagctctcaagctccagctgttcgct | 3060 |
| Db | 2760 | gggaattcggtcagattatccgcacacgatggccgtgattcttcaaaaggccacattatagt | 2819 |
| QY | 3061 | gccagaaacagtcagctcaataatcattgctccacatacgaaggtctccatcagtaga | 3120 |
| Db | 2820 | gccacacacagtcattgctcattcaaacattgctccacatcaggctctcttatcgagtaga | 2879 |
| QY | 3121 | tgttaagaagaaaggtlcatgtgcgctagtcgcccctccctcccaagaagatgtga | 3180 |
| Db | 2880 | tgttataggaagactaggtcattgctcattgctccacatcaggctctcttatcgagtaga | 2939 |
| QY | 3181 | acaggggcacccgagggacagctgtctactctgccaatacaacaataagcccaatgatag | 3240 |
| Db | 2940 | acagggcgcttcacaggagacagctgctactcctgccaatacaataatgacggatggacag | 2999 |
| QY | 3241 | acacgctlcatagagtgaccactactctccagaataagacagagattgtgaagcagaga | 3300 |
| Db | 3000 | acacggctgcatggacgacacattactcttcagagaaagacagtcattttctcacttacc | 3059 |
| QY | 3301 | tagacagccata tcaacagatccagatcaacagacacacgcccctccctagacgagcac | 3360 |
| Db | 3060 | tgcgctctgcacacagacacacagatgagatcaccacagagat | 3102 |
| QY | 3361 | caaccgctccagatccctctgaaagcttcgatacaaacctcaagtgatgcctcatt | 3420 |
| Db | 3103 | ----- | 3102 |
| QY | 3421 | aatgactggaagatctgcccctccctccactcgtatcgaggtctcaacctgtaccgt | 3480 |
| Db | 3103 | -----GGAAGGTCTTACCCTCGCACTGG | 3125 |
| QY | 3481 | gtctgtccagacaagccca tcaagtaactccggaacagacgagggccgcacagcttc | 3540 |
| Db | 3126 | ctctgtccagacacattcctctcaagtaactccgtaacacagacgagggccgcagcttcc | 3185 |
| QY | 3541 | acagcttccaccaaagggacacttggagagaaagtgtcatgtatagatagagagaaatcg | 3600 |
| Db | 3186 | TCAGGTCCCAACAAAGGGAACACTGGAAAG | 3215 |
| QY | 3601 | ccaattgaacttaacaatatacaacagtgatgcggatcaagacccacagactygaacaga | 3660 |
| Db | 3216 | ----- | 3215 |
| QY | 3661 | ttaccattcgaaaglatcgctcagatggatccacatatagaggggcagatactgttccac | 3720 |
| Db | 3216 | ----- | 3215 |
| QY | 3721 | taaatccctcgcagatglatgtlaatgtatctctgcggttccaagagatagtagtcttc | 3780 |
| Db | 3216 | ----- | 3215 |
| QY | 3781 | tcgttccagcagacacaagctacatgtcgtlccaatcaagacggccgagagagaaacagaa | 3840 |
| Db | 3216 | ----- | 3215 |
| QY | 3841 | aatcagtgctttacatccaaaatgcaaaacagacagatggcggtgtccgggaaagactt | 3900 |
| Db | 3216 | ----- | 3215 |
| QY | 3901 | gaaccaaaagcaccagcatcagtgtgagacatgtgtcactctgaaagaagaatgacggcagca | 3960 |
| Db | 3216 | ----- | 3215 |
| QY | 3961 | gtccacacactgacgtggggcccttggttacccagtggcagaagaacgggcagactgcatltyg | 4020 |
| Db | 3216 | ----- | 3215 |
| QY | 4021 | ggccaaaatgtagctattgttctgtctcaccgaaaagtgcgagtgctcctcaactcag | 4080 |
| Db | 3216 | ----- | 3215 |
| QY | 4081 | ccaacccgaagagaggttaaaaagctacgagacactgttcaagagaacacagagaccg | 4140 |

[illegible]

| | | | |
|---|------|--|------|
| D | 1569 | AAGCGCTAAAGATGAGAGTGTACTCGAGATTTCAGAGACAAATGCTTGCTTAAGTT | 1628 |
| O | 2108 | gtagaaggaagaatgacgtgaatcgaagtcgaacttgcatttattacaaagtaaaaaa | 2167 |
| D | 1629 | GTAGGCGAAGAAATGACTGAATCAGCTCGGCTTGTGCATTATTACTAAAGTAAAAA | 1688 |
| O | 2168 | ggaatttgatgatactctgtagacactttagaccaggtgataagctcttggaatgaa | 2227 |
| D | 1689 | GGAAAGTTTAGTGATCTGTAGGACATCTTAGACACAGGTATGAAGTATTAAGAAATG | 1748 |
| O | 2228 | ggagagctatctcaagagccacatttgagaagttacaacattacagaatccaa | 2287 |
| D | 1749 | GGAAAGCTACTCGCAAGAGACCCACATTTGAGAACTGTACAAACATCATTTAGAAATCCAA | 1808 |
| O | 2288 | ccctgaaccaagaattgagcttgttgcataaagccaattgagataatctcctaagaa | 2347 |
| D | 1809 | CGTGAACCCAGATGAGATCTGTAGTTTCAGGCGCTATGAGATATACCGCAATACCT | 1868 |
| O | 2348 | gatagcacgcatagcacaacttggaaatccagttctagctcatttgaatctcaaaaaatgac | 2407 |
| D | 1869 | GATAGACACATGACACACTGAGTCAGTCTAGCTCTTTGAATCTCAAAAAATGAT | 1928 |
| O | 2408 | cgctcctctatatccgttaacctaccatgagtcctgagctgagagatgtcccgag | 2467 |
| D | 1929 | CGTCTCTATTTCTGTACCTCTCCCATGAGTCCTGGAATGTTGAGGAGATGCCACAG | 1988 |
| O | 2468 | ttcctatctgcagacgtcttcaataaactatggttgcagaagttgtgtccagttgata | 2527 |
| D | 1989 | TTCTTATCTAGGACCAACTTTCATTAATAAACTATGCTTGACAAAGTTGGTCAACATTTATA | 2048 |
| O | 2528 | gttaaatcttggagcaagaatctccctccagggaaatgtggagccaagaatccct | 2587 |
| D | 2049 | GTTCACATTTTGGGACCAAAAGATCTCCCTTCAGGGAAATGGAGGCCAAAGAAATCCT | 2108 |
| O | 2588 | taqttaaatattactctctccagataagaatgataaataagaagaacaaacaa | 2647 |
| D | 2109 | TATGTTAAATTTACTCTTCTCCAGACAGAAAGTATTAATAAACAAGAACACTAAAAACA | 2168 |
| O | 2648 | gtcaagaanaacttggaaacccaaatggaacagacttcaattctcctgtccacga | 2707 |
| D | 2169 | GTAAAGAAAACATTTGGAAACCAATGGAACCAAAACATTTATCTCCAGTCCACCGA | 2228 |
| O | 2708 | agaagaatccgtaagaagaatgtcgaataatcaccttggatagaagtacgttcgaa | 2767 |
| D | 2229 | AAGAAATTTGGGAAAGAAAGCTAGAGATTACCTTTGGATCAAGCTCGGTCGAGAG | 2288 |
| O | 2768 | gaagaagcgaatctcttagagaagatttaattgaattggaacagcttgcagatgat | 2827 |
| D | 2289 | GAAAGAAAGTAATTTCTTAGCGGAGATTTTAATTGAATTGAACAGCATTAATTAGATGAT | 2348 |
| O | 2828 | gagccgactgtgataagctgcagaccatgatagtctcctcattgcacactccctcgccct | 2887 |
| D | 2349 | GAGCCACATTTGGTACAAACTCAGACGCATGATGTCTTCAATGTCACATTCGCCACCT | 2408 |
| O | 2888 | tcccatatctcccccgaggagcgtccatctgagagagcccaacgcgcgaagctgcagaag | 2947 |
| D | 2409 | TCTCATATATGCGCAGGAAGACGCTCCATGAGAGAGCCCAACGAGGATTCAGAAAG | 2468 |
| O | 2948 | tcgaagaagaatagtcagagtgctgcatacagctgcagagatggcgltggagta | 3007 |
| D | 2469 | TCAAGAGAAATAGTATAGTGAAGTCTGTGACTATGCTGATGATGGAATTTGCTGTA | 2528 |
| O | 3008 | gtgtcaagattatcgacaacaatggcgcgactctcaaaagctccacgltgtcgtgtccagaa | 3067 |
| D | 2529 | GTATAGATTATTCGACATATGCTGAGATCTTCAAAAGCTCAACATTAATCAGTCCAGAA | 2588 |
| O | 3068 | caagcatgtcatcaataatgctcagatcaggtctcctcatcagatagatgata | 3127 |
| D | 2589 | CAGGTAATTTCAATCAACACACTGTTCAACATCAGGGGCTCTCTCATCGATGAGATGTTATA | 2648 |
| O | 3128 | ggaaggaacaaggtcattgtgcgcttagtgccccctccctcaaa---ggaatgtgaaacag | 3184 |

| | | | |
|---|------|--|------|
| D | 2649 | GGAAGACTAGATCATGTACACCCAGTGTCCCTCCACAAAGTCCGAATGTGAACAG | 2708 |
| O | 3185 | gggcaaccgaagggaacgylct---actgccaattacaacaacttagccgaatgataga | 3241 |
| D | 2709 | GGGCTTCGAGGGACCCGCACTATGACCGGACGACATTAATCAATTACCCGATGAGAGA | 2768 |
| O | 3242 | caacgtatcatgtatgataccaactctctcaagataagagacagggatlttgagacagagat | 3301 |
| D | 2769 | CATCGTGTATGGATGATCAATTAATTTCTCCAGATAGACAGGAGTTGTGAAGCGAGAT | 2828 |
| O | 3302 | agagaccatatacagagatccagatcaacagaaacagcgctctcctagagcgagaccac | 3361 |
| D | 2829 | AGACAGCCATATCAGATGCCATGCATCAACAGAAACAGGCGCTCTCTGTAGCGGACCAAC | 2888 |
| O | 3362 | accgcgtccagatcctctgaacgltccctgatacaaacctcatalgagltcagltcattla | 3421 |
| D | 2889 | ACCCGCTCAGATCCACTGATCAACGCTCTGATACAAACCTCATGAGTGTGATCCCTTCA | 2948 |
| O | 3422 | atgactggagaatctgccccctccttcaacctgacttatacgaggtctcaacctcgtacggg | 3481 |
| D | 2949 | ATGACTGGAAGATCTGCCCTCTTCACTGCTTATTCGAGGTCTCATCTGTACTGGG | 3008 |
| O | 3482 | tcgtccagacaagcccatcaagtaactccgaggaacagagcaaggcgacagactcca | 3541 |
| D | 3009 | TCTGTCCACACAAGCCCATCAATCACTACTCCAGTGCAGAGAGGAAGGGCCGACACTTCCA | 3068 |
| O | 3542 | caagctccacaagaagggaacatltgagagaaatgtctatgtatataagaagaaatcgc | 3601 |
| D | 3069 | CAGCTTCCACCAAGGAGAACGTTGTGATAGAA----- | 3099 |
| O | 3602 | caaatgaacttaacaataacaacaggtagccggtatcaaccccaactggaacagat | 3661 |
| D | 3100 | ----- | 3099 |
| O | 3662 | taacatcgaagtatcgtccagatgtgataccataagagggcagatactgttccact | 3721 |
| D | 3100 | ----- | 3099 |
| O | 3722 | aaatcctcggaagatgataagatgtatctgcglttcaagactagtagtgcctc | 3781 |
| D | 3100 | ----- | 3099 |
| O | 3782 | cglttcagagcaacaagctcaatgtctgcctcaatcagagagcgccgagagaacagaa | 3841 |
| D | 3100 | ----- | 3099 |
| O | 3842 | atcacgtcttataccataaaltgcaaaacagacagatggcgltgtcgggaagaacttg | 3901 |
| D | 3100 | ----- | 3099 |
| O | 3902 | accaaaagcacaagatcagltgagacatgtgtcaactggagaagaatgaagcagcag | 3961 |
| D | 3100 | ----- | 3099 |
| O | 3962 | tccgacactgcagtggtggccctcggttaccagatgcaagaagcggcgatactagatggg | 4021 |
| D | 3100 | ----- | 3099 |
| O | 4022 | gccaatatgtagctatgttgtctctcaggaagaatcgcsagtgtcctcaactcagc | 4081 |
| D | 3100 | ----- | 3099 |
| O | 4082 | caaacggaaggaggtataaaaagctacggagcactgttcaagaagaacagcgagccggg | 4141 |
| D | 3100 | -----AAGCAGGAGGTAAAAAACTAAGAGAGCATGTCTCCAAAGAAATTAAGAAACAGGC | 3152 |
| O | 4142 | ctagcaatgtgaatgagagactgatacccgccagcgccgagcgagataccaagaatggc | 4201 |
| D | 3153 | CTGGCGGTGGAATGAGGATGATGATCTGCAAGGCAAGCCGAGGTATACAGATGGT | 3212 |
| O | 4202 | agcatgaacagatataagcttcggaagaatctgatctctcccttaggtgtccgcctgcgctc | 4261 |
| D | 3213 | AGCATGAACAGCTACAGCTCAGAGGAATCTGATTTTCCCTGTGTGCTTGCCCTTGCTCT | 3272 |

| | | | |
|----|------|--|------|
| QY | 475 | tttcattatgccaaaccaagttctctgtgctcgatgltgtagtgcagtgctcttaagctcaaa | 534 |
| Db | 879 | CTCCATTATTTTGTCCACCAAGTTCTGTGTCACGCTGCGGAGGCCGGTGTCTCTGTGGATTGAA | 938 |
| QY | 535 | caae-----gattatgtgtgtgtgtatattgtgtccgaaacaagaatccct | 582 |
| Db | 939 | CAATGAGCAACAAAGTGGTTATGTGGTTATGCAATTTATGTCCGAAGCAACAGAAATCTT | 998 |
| QY | 583 | cactaatcagtagagcatggttataatagtggtgtctaacacatgcagcaacctgtca | 642 |
| Db | 999 | AACGAAATCTGGAGCGGTGGTTCTTTGGAGTGGCCCT-----CAGCAGCCTTACTCA | 1049 |
| QY | 643 | aaaggttctctc-gaagggcttcgaatgagaagccctctgaaggaagaagcaaaactac | 701 |
| Db | 1050 | AGATGGACTGTGAGTGCACGCGCCACAGGTGCTGATCTGAGGTGCCAAGAGAAAGAA | 1109 |
| QY | 702 | acgagcaagcccccagttccaaagagagcccccagttgactatcagtaacctgcagttgaag | 761 |
| Db | 1110 | AGCAAGGCTCCAAAGAGCGATTCAAAGGTTCTGACAGCCCTTGATGATACAGCACTGTCCTTC | 1169 |
| QY | 762 | gcccagcttcattggtgtctacaagaacaggtactattaaatgtatcagaagtgaagcacc | 821 |
| Db | 1170 | CCAAAGACACTGCTACCCCCCGGTGTCACCGCTTGCCACAGAAAMAAAGGGCTGAGCCCTTACA | 1229 |
| QY | 822 | agattgcagttacatgacctctcagacagaaagaagttccatcagtgltccagtgatcaaa | 881 |
| Db | 1230 | GCAACCTTTGGTCTCTGTAACAGACAGCAGCATCAAGATCAAGAAAGCAGCCACCGAGGA | 1289 |
| QY | 882 | atcgaagatcagagcaaaagtgaagaagaagagactactacagatglttccttcagatg | 941 |
| Db | 1290 | AAGGAGAAAGGCTCCAGGGCTTTTCAGAG-----CAGATGGAAGGAGGAGGCGACAGAG | 1341 |
| QY | 942 | gtacatatcccaagatctctcttcagatgcatgcatgatacagatctcagccgtagactaat | 1001 |
| Db | 1342 | AGCGAGCCCAACGTTCTCCCAAGTTGTGTGTCAACCCCGGGAAGGAGTGCAGATGAG | 1401 |
| QY | 1002 | tttatgaagaacctgtgcattttaattacaagggattcctaagagagagccatagacatt | 1061 |
| Db | 1402 | AGGGAGAGAAAGAGAGGGGGGGAACCCGAGGTTGGAGAAAGGGGCGCTCCACAGACTAC | 1461 |
| QY | 1062 | ccaaagagtatcttgtgattgtatgaagaatgltgtagagcagagatgtaatlgaagaagcaaa | 1121 |
| Db | 1462 | TCAGACCCGGCCT-----GAGAAACCGCAATGTGCAGGGTGTGGGGAAGACCCAGAGGAGA | 1516 |
| QY | 1122 | ggaagagaggaatataccgcgacgcgtacagaagtbatccaatctgtcccggtatccg | 1181 |
| Db | 1517 | GGAAGAGGAGGAGTAGTACGACTAGTACCGCAGCCACTTAACCTGTGGTGTGTAACCCG | 1576 |
| QY | 1182 | taagccacaaccttacgaagaacaaatgcgcaltccaagcttgtagtgtlccagggcacagc | 1241 |
| Db | 1577 | TGAAGGCGCGCCGACAGGAGCAGCAATGTGCGCATGTGACGCGCGGTGTCCCGACGAGGC | 1636 |
| QY | 1242 | atgagaagaagacagatgtatgtttcttggcaagcgttgacatagaagaattccagatt | 1301 |
| Db | 1637 | ACGAGCGGCGCCACACCGCAGTGTGGCTTCCGCACACCGAGGCAAGCTGCCGCCGCCGCG | 1696 |
| QY | 1302 | ctctgtctaagatgtagatgacacatcaagaacaaagatctgtatctgtaacgttagagctgca | 1361 |
| Db | 1697 | CTGAGGCGCACGGCGGGCCAAAGCGCGGCCGCCA-----CCGGCA | 1735 |
| QY | 1362 | tggaataaccacagatcgttatctcaatgtgaaagaactcgagaagctcagtagaacaagtctc | 1421 |
| Db | 1736 | GGGTCTCTCCCGGAGTCCCGCGCCGCAAGCCCGCGGCGCGCCACACTTC----- | 1786 |
| QY | 1422 | atccacaagaagcttcaaatcatatagttcctccaagcccttgcgagagacccataacagcttg | 1481 |
| Db | 1787 | -----CCACCGAGCAAGGGGCCAACCGCGCGGCCACGCCCGGGTCTCCGAG | 1833 |
| QY | 1482 | atagaccagacatgagcgcgctgactcctctacgcgaagaacagcaaccttagatccagct | 1541 |
| Db | 1835 | AGCCACCCGAGCGCGCGCTCTCCCGACCGCTGCTGTAGCAGGGCGCGCTGTGACCCGGCT | 1894 |

[illegible]

Db 2825 GAAAGCAAAAGTAAGAGAAACCAAAACAGTAAGAAACTTCTAGAGCCAAAATGGA 2884
OY 2676 accagacttcatltaatctcctgctcaccggaagaattccgttgaacgaattgctgaaa 2735
Db 2885 ACCAGACATTGTTGCTACTACAGCTACATGTAAGATTTTCGAGACGAATGTTAGGA 2944
OY 2736 ttacccttggatcaagctagagttcgcgaagaagaagacgaattccttagagagattt 2795
Db 2945 TTACCGTGTGGGACCGAGAGATGACAGACGAAGAGATGTAATTTCTTGGAGAGATCC 3004
OY 2796 ttattgaattggaanaagcttgccttagatgatgacgcgaactggtatagaattgagaacc 2855
Db 3005 TCAATTAGTTGGAAACAGGCTTTTATGATGATGAGCCCAATTGGTATTAACCTCCAGACAC 3064
OY 2856 atgatgtccctcatatgtccactccctccctccctccatctgccccggagagcagctcc 2915
Db 3065 ATGACGAATCTTCACTACCTCTGCTCAGCCATCACCGTTTCATGCTCCAGGGGCATATTC 3124
OY 2916 atggagagagcccaacgcgcagctgcaaaagtgcgaagaataatgtacagtgaaagtgt 2975
Db 3125 ATGGAGAGAGCTCCACAAAACCTAACAAAGATCTCAGCAATCAGTATGATGATCATCT 3184
OY 2976 ctgacacagactgcgagagatgcgcgtgagtgatgtgc-----agattatgcacaaatg 3029
Db 3185 CAGATTATAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3244
OY 3030 gcgcgacttccaaagctcagctgtgcgtgcgcgaagaagtcagtcagtcagtcagtcagtc 3089
Db 3245 CTAGAGAGAGTAAAGAACCAACGTTAACTGACGAGCAAAAGAACTACATCAACC 3304
OY 3090 gctacacatcaaggtctctcctcctcctcctcctcctcctcctcctcctcctcctcctc 3149
Db 3305 GCTCACGTTCCGCTGCTCCTCATCGCGCATGATCAGGGAAGGCTCGTTCACGTTTAC 3364
OY 3150 ctgagtcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 3206
Db 3365 CAAATGTGCTTACAGAGAGAGCTTACATGTAATTCATCCAAACGAAAGTCCAGCTTTC 3424
OY 3207 ctgagccattacaacacaaatlagccgaa-----tgatagacacgcgtgctatgagtg 3257
Db 3425 CAACCCGACACCATGATGCTCCGGAAGCCCGCGCATCACAGATCCAGACATTTGGA 3484
OY 3258 accacactactttagatagagacagga-----ttggaagcagcagata 3302
Db 3485 GTCAATATTCTGTAGAGCCGACAGTAGTCTTCATGCTGCCAGACAAACGAGGAC 3544
OY 3303 gacagccataatcagatccagatcaacagaacagcgcctcctcctcctcctcctcctcctc 3362
Db 3545 GAAGTGCAGAAAGCTTACATGATGACAGCTGACAGCCCTCTTGCACAGGGTTAGGA 3604
OY 3363 cccgctccagatcctctgaaagctcgtatatacaaaactcctaagtgagtcg-----atgc 3413
Db 3605 GTGCTAGTACCAATGCTGTAGACCAAGATACTAGTTTCATTCACCAAGAGAAAGGC 3664
OY 3414 cttcattatagactggaagttgcctc-----cttcaactg 3452
Db 3665 ACTCCAGAAAGTCTGAAAGATGATGATCCAAAACAGCTTAGGAAGAGCACAGCTCTG 3724
OY 3453 ccttactgagttcaccctcgtacccggtcgtcgcagacaagcccatcaagttaccg 3512
Db 3725 ATGCAACACGAAGACCGACGAGAAAGGCCAACCCAGTCTCTCTCCAGCAGACACATCT 3784
OY 3513 gaacagagcgaagggcgacagcttccacagcttcca-----ccaagggaaatgagga 3569
Db 3785 TCGGCGAGTCCGTGGAAGACAGCTCCAGAGTCCGAGTTCGAAGCGCGAGTATAGAAC 3844
OY 3570 gaagtctatagatatagagagagaatcgccaaatgaaactaa-----ca 3617
Db 3845 AAGCAAGCTTAGTAGTAGAGAGGAAAGAGACAGATGAAGTGAAGTTCAACCGATTTA 3904
OY 3618 aatacaaacaggttagccggtatcagaccacagactggaagcaagattaccatcgaagtac 3677

Db 3905 AGCAGCAACAGGGTCTGGGTCTAGTCAAGAACTTGACCAAGCAAGCAATACTCCAGTACA 3964
OY 3678 gctcagatggtatccacatagaggggcagatactgtttccatlaattcctcgtgagtg 3737
Db 3965 ACATACATAAAGATCAGTACAGAAAGCTGTATACCCGCTGTGCAAGTCTTTCAGATAGTG 4024
OY 3738 atgtatgatatatctgagttcaggaactagtagtctgtccttccatgagacaa 3797
Db 4025 ATGTCAATGATGTGTCCGCAATTTCCAGAGCCAGCAGTACTTACGCTTACGACGACAA 4084
OY 3798 gctacatgtccctccaatcagagcgccgagaggaacaggaatcagtgcttccat 3857
Db 4085 GCTTATGTCAGAGCAGCTGTGAGCGCCCAAGG-----GTAGATCATGTTATTACCC 4138
OY 3858 ccaaaatgcanaacagacagatggtgcgtgtcggggagaacttgaccaaagaaccagca 3917
Db 4139 CCAAAATGCAGACAGACAGATGCGGACCTTCAGAGAGAGCATCATMAAGACACAGTG 4198
OY 3918 tcaatgagacatgtgtcactactgagagaagaaatgacgcgaagtcagactgagtg 3977
Db 4199 TAAGTGAGAGATATATACCTTGAGACGTATATACGCTTACCGACTCGACACGGCGTAG 4258
OY 3978 gcgccttggtagccagtggaagaagcgcgcatcagatctggtggccaaatgtagta 4037
Db 4259 GTACGTCGAGGCGGTGGAAGAAAGAAAGAAATCCAGCTTGAGGCGCAAGTGGTAGCA 4318
OY 4038 ttgtgtgtctcacaaggaagtcgcagtgctcctcaactcagccaacccgaagaagag 4097
Db 4319 TTTG-----TGCTCCAAAGAACGAGAGCAGTACAGCTCAGCACAACAGTGGGCC 4372
OY 4098 gtaaaagctagagagcactgttcagagaagcagcgagaccggtcagcgtgagatga 4157
Db 4373 ACAGAGATTGAAGAACCAACCATCCAGACGATACGGAACAGCAATGCGACGTCAATGC 4432
OY 4158 ggaactgtgatgaccgcgcagcgccagcggaatccacagatggaagatgaacagctata 4217
Db 4433 GGAA---GATGAGAGACAGCCGAGCGGAGTCCAGATGCGAGATGCGAGATCAACAGTTATA 4489
OY 4218 gctcgaagaagaatcgtatctcctcctcctcctcctcctcctcctcctcctcctcctc 4277
Db 4490 GCTCGGAAGAACTTGATATTTCTGAGTTGAGTAGGACCGGACAGTCAAGTTCAGTG 4519
OY 4278 atttccgtatgctcgtgggcctcgtcagctagttggaacgcgaacctgtgctactctg 4337
Db 4550 ATTTCCTTGAATGGGTGGGACACAGCGAGCTCGTTGGCCCTCAGACCTCGCCACCCGG 4609
OY 4338 caatggtgacattcagtggtgaatgatgtataaagaagacagcttgaggttagaataca 4397
Db 4610 CCATGGCGCATATCCAAATCGGATGGAGATGAAGAGGGTCAAGTTGAGGTTA 4469
OY 4398 tccgggcgcgcgccttgtgtataaaccaagttccaaagacacgcgcgaaccggtatgca 4457
Db 4670 TCAGAGCCCGGAGCCCTTACCAAAACCTGCTTCAATCTTACACCCGCTCTATGTGA 4729
OY 4458 aggtgatctgttgaacaagagttgcatagccaaagaagaacccaagtggtgcgaa 4517
Db 4730 AAGTATATCTTTTGGAAATAGAGGCTGTATTCGCAAAAAGAAAGACAAAGATTSCAGGA 4789
OY 4518 agaccctgagccctgtacacagactctgtcctcctcctcctcctcctcctcctcctcctc 4577
Db 4790 AAACCTCTCATCTTGTATTCACAGTCCCTGGTTTGTATGAAGATCCACAGGTTAAG 4849
OY 4578 tgttaagatcatctgtcctgggagatataatgtcgtatgatatcaaaatccctttagggag 4637
Db 4850 TTTCTAGGATGATGTGTGGGTGACTATGTAAGAAATGACCAAAATGCTTATGAGGTG 4909
OY 4638 tggccagatcactcttagatgaaactaactaactaactaactaactaactaactaactaact 4697
Db 4910 TGCTCAAAATCTTGTGGAAGAACTTATCTATCCACATGTGATGATGATGATGATGATGAT 4969
OY 4698 tcttccctcctcctcctcctagtagatcaaacctcgcactctgacaagaagagctccc 4757
Db 4970 TGTTCCTCTCCTCCTCCTCCTGATGCCACTGCTCCCTCCGACCCGCGGCTTCC 5029

1062 ccaaaaglatatctgtgtagtgaagatctgtagagcgcaagatgaatatgaaagacaa 1121
1462 TCAGACCGGCC-----TGAGAAACGCGACAAATGGCAGGGTGGGGAAGACCGAGACAGA 1516
0Y 1122 ggaagagaggaataacagagcagctacagaagtgaatccaaatctgcccgtatccg 1181
Db 1517 GGAAGGAGGAGGACTACCGACTAGTACCGGACGCACTTAACCTGGCTCGTACC CGG 1576
0Y 1182 taaagccaaacctacgaagaaacaaatgcgcatccaagctgaagtgctccaaggaagac 1241
Db 1577 TCAGAGCGCGCCGAGAGGACACAGATGGCATGGACGCCGGGTGTCGCGAGAGC 1636
0Y 1242 atgagagagagcgaatgtagtcttcttgccaaagctgaactagaagtccaagattt 1301
Db 1637 ACGAGCGCGCCACACGAGAGTGGCGCTCCCGCACACGAGGCAAGCTGCCCGCGCGG 1696
0Y 1302 ctctgctaagatgtagtagaccatcaaggaagaatctgtatctgaacgtagaagctgca 1361
Db 1697 CTGAGGCGACGCGGCGGACGCGCGCGCGCA-----CCGCA 1735
0Y 1362 tggaaacccaagatgtagtataatggaagaactcgagaggtcgaaggaagattctt 1421
Db 1736 GGGTCTCTCCCGGAGATCCCGCGCACGCGCGCGCGCCGACGCTC----- 1786
0Y 1422 atccaaagaagacctcaatcatatagctccaccacctcgcgagacctataccgcttg 1481
Db 1787 -----CCACGAGACGCGGCGCACGCGCGCGCGCGCGCGCTCCGCGAG 1834
0Y 1482 atagaccagaatagagcgcgctgaactccctacggaacacgacccaacttagtccagtt 1541
Db 1835 AGCACCCGAGCCCGCGCTCCCGAGCGCTCCGTAACGACGCGCGCTGGACCGCGGCT 1894
0Y 1542 ctgc-----tgtagaagaaacggaagaaataatgaacacatgttaagaatgatt 1595
Db 1895 CGGCGGTCTCTCGGAGAGGCCAAGCCGAGAAAGCGGAGAGCATGCTGGGAAAGACT 1954
0Y 1596 cttagtctacagacagctcagagtcagtcagtcgccccaccacgaagcctcataatcca 1655
Db 1955 CGCTGAGTCCGATCAGTCCGAGTCCGTGGCGCATCCCGCCCAAGCTCAACCGGCCA 2014
0Y 1656 agaaagaggtaaatgtgcgaagtttcaactgaagacgtcgagagagagcgtagtcca 1715
Db 2015 AGCGGCGGAGAGGAGAGCTGATGTCGTGAGAGAGCTCGAGAGAGAGGGGCTGTCCA 2074
0Y 1716 cactgtagtatcaagctgtgtagtctgtagctggaagcgagagtgtagtgaagaaag 1775
Db 2075 CACCGAGATACAGAGAGCTGCGAGAGGTGAGCTGGAGAGCGAGAGCGGAGAAAG 2134
0Y 1776 gggacagtcacaaaggaagaaataactgtagagagaggttcttgtagtcttaaca 1835
Db 2135 GTGACTT----- 2141
0Y 1836 ccaagctcgaagacacaaagaagatgtactatgtgccaactcttllgaaagagatt 1895
Db 2142 -----GATTT 2146
0Y 1896 tggaaatgtctgagcctcagatgaagactctggttagatcactgttagcaaaccc 1955
Db 2147 ACTACTGGTTGATCCCGCACGTCGACAGGAGGAAAGGTGCCCTATAG----- 2198
0Y 1956 ttaacagagagatagcctatagtagtaagaacacctgtagctgcaagcctccaagattg 2015
Db 2199 -----TTGCATCTCTGTAACTGTGACCGCTTAAAGAGG 2233
0Y 2016 gagatcgactaatgtgctgattatlaataaagcgtttaaagaatgtaggtctac 2075
Db 2234 GAGATCGATTAATCGGCCGTTATTCTTAACAAAGAACAAAC-----CATGCCCA 2284
0Y 2076 gagaactcaagcaatgtcgtggtcttaaggttctgtagaagaaagatgaactgaacaggtc 2135
Db 2285 AAGAAATCAGGTGATTAATGGGTCTGAAGGTGTTGAGGAGAAATGACGGAATTAGGGC 2344

0Y 2136 gacttctgcatattacccaagttaaaaaaggaagtttaagctgatactgtagacatc 2195
Db 2345 GCGTTGGTGTCTTTCATCAACCAAGTAAGAGAGGCGCCTGGCAGAGCTGCGACACC 2404
0Y 2136 ttagaacaggtatgaagtctctggaatggaaatggagagctatgtccaaggaacatttg 2255
Db 2405 TAAGAGCAGGGAGCAGAGCTCTAGAGTGGAAATGGTAACCCCTCGCGGAGCAAAACG 2464
0Y 2256 aggaagttacacattatctagaatccaacacctgaacccaagaagttagctgtgttt 2315
Db 2465 AAGAAATTTCACAACTTATCTTAGAATCAAAATCAAACTCAAGTTGAATTTATTTGTTT 2524
0Y 2316 caagggcaattgagatattcctagatattcctgatalagcagcatgcaacatgaaatcca 2375
Db 2525 CAAGGCTATTGTTGACATCCCGAGATCCCTGAGATTCCATTCCTCCCTGGAGTCCA 2584
0Y 2376 gttctagctcatttgaatcctcaaaaaatgagacgctcctctatataccgttaccaccca 2435
Db 2585 GTTCAAGTTCCTTGAATCTCAAGAAATGGAAAGCGCTTCTATTCTGTATTTCTCCAA 2644
0Y 2436 tgaagctcgtgcatgtcgaagagatgctccgagttcctatctggaacgcttcaataaac 2495
Db 2645 CCAGCCCTGAGCTCTGAAGATGGCCCAAGTCTTACAGGCGCACTCTCAGTGAAGC 2704
0Y 2496 tatgtttgacaaggtgtgacacagltgaatgataatttggagcaagatctcc 2555
Db 2705 TATGATATGATTAAGTGGGCGACACAGCTGATTGTAATGTCTCAAGCAACAGATTCAC 2764
0Y 2556 ctcccaaggaagatggagagccaaagaaatcctatgttaagattactctctccatagata 2615
Db 2765 CCGCTAGAGTAGATGGCGCGGCCGAGGAAATCCCTATGTAAATGTATTTTCCAGATA 2824
0Y 2616 gaaggttaaaatgaagagagaaacaaacagtcacaaacacttggaaaccaaayga 2675
Db 2825 GAAGCGCAAAAGTAAAGAGAAACCAAAACAGTAAACAACTCTAGACCAAAATGGA 2884
0Y 2676 accagacttcaattatctctcgtlccacggaagaaatccgtgaaagaaatgtcggaaa 2735
Db 2885 ACCGATATTGTGTACTACACAGTACATCGTAGAATTTTCGAGAGGAAAGTTAGAA 2944
0Y 2736 ttacccttggatcaagctgaagattctgagaagaagagcgaaatctttagagagatt 2795
Db 2945 TTACCGTGTGGACCGAGAGAGTACAGAGCAAGAGATGAATTTCTTGAGAGATCC 3004
0Y 2796 taattgaattgaaacagcttctgtagatgtagtagcgcaactgtgataagctgcagacc 2855
Db 3005 TCATAGAGTTGAAACACGCGTTTATGATGATGAGAGCCCATTTGTATTAACCTCAAGAC 3064
0Y 2856 atgagtctcctcatgtccatcctctgcgccttcccatatctgccccggagcgaaatcc 2915
Db 3065 ATGACGAATCTTCACTACTCTGCTCAGCCATCACCGTTTCATGCCAGCGCGCATATTC 3124
0Y 2916 atgagagagcccaacgycgcgaagctgcgaaggtcgaagaaatgaatgaagtgaattg 2975
Db 3125 ATGGAGAGAGCTCCAGCAAAAGCTAACAAAGATCTCAGCGAATCAAGTTAGATCATCT 3184
0Y 2976 ctgactacgactgcgaagatggtcgtggaatgtgtc-----agattcgacaaatg 3029
Db 3185 CAGATTATGAGGTTGATGATGTGATGAGTGTGCTCCAGTGGGTTATAGAGCTAGTG 3244
0Y 3030 gcccgatcttcaagatcctcaagctgtcggtgcgaacaaacatgcatgtcaatcaatc 3089
Db 3245 CTAGAGAGAGTAAAGCAACACGTTTAACAGTGGCAGAGCAACAAAGATCAACATCACC 3304
0Y 3090 gtccacacatcaaggtctcctcatcgaatagatgtltaagaagagaacagatcagtgtcgc 3149
Db 3305 GCTCAAGTTCGCGTCTCTCATCGCGGATGATGAGGAAAGCCCTCGTTCAAGTTTAC 3364
0Y 3150 ctagtgcctctcctctcaaggaatgtgaa---caaggcgcaaccggaggaacgtgtcta 3206
Db 3365 CAATGTGCCATTACAGAGAGCTTAGATGAATTCATCCAACACGAAAGGTCACTCTC 3424
0Y 3207 ctggcattacacacaaatlagccgaa-----tgaatagaacacgtgtcatgattg 3257

```

Db 3425 CAACCCGACACCATGATGCTCCCGAAAGCCGCGCATCCACAGATCCAGACATGTGGAAA 3484
Oy 3258 accactactcttcagatagagcaagg----- 3285
Db 3485 GTCAATATTTCGTGAGAGCCGACAGAGTACGCTTCTCATGCTGCCAGAGCAAAACGAGGAC 3544
Oy 3286 ----- 3285
Db 3545 GAAGTGCAGAAAGCCTACACATGACCACTGACAGCCCTCTTTGACAGGGCTAGGA 3604
Oy 3286 ----- 3300
Db 3605 GTGTAGTACCAACTGCTTGAGACAGATACATGTTGATTCACAGACGAGAAAGGC 3664
Oy 3301 ----- 3300
Db 3665 ACTCCAGAAAGTCTGAAAGATGAGCATCCAAAACAGTCTAGGAAAGGACACAGCCCTCTG 3724
Oy 3301 ----- 3329
Db 3725 ATGCAGACAGGGTTCTCCACCATGCTTTCTAGAAAGGGATACGCAACCCCAAGAGCAA 3784
Oy 3330 cagaaacaagcgctctcctagagcggacacacacccgctccagatcctctgaagctctg 3389
Db 3785 CCGATCAACCGGTGCTTAGGGGAAAGTATCCCACTGCTACGCTGACGGGAGCAGCTCTA 3844
Oy 3390 atacaacctcatgagtgctgagtgctcattcaatgaatgaatctgcctcctc--- 3446
Db 3845 GTGTGAGAACCTGTGTTCTTATGACACACCTTGCCCGCGAGGGTCCGCGCCACTTCTC 3904
Oy 3447 caactgccttatcgaagtgctcaacctcgtacccggtctgcagacagaagcccatcaagta 3506
Db 3905 CACTTCTTTGACAGACAGCACCGACAGAAAGCCCAACCCAGTCTCTCCACAGACAA 3964
Oy 3507 ctccgggaacaggaagcgagcgacagctccacagctcca---ccaaaggaaacat 3563
Db 3965 CATCTCTGGCAGTCCGCCGTGGAGAGACAGCTCCACAGGTGCCAGTTCGAAAGCGGAGTA 4024
Oy 3564 tggagagaagtgctatgatatagagagagaatacgcgaatgaactaa----- 3615
Db 4025 TTAGACAAAGCAAGCTTAGTAGTGAGGAGAGGAGCAAGACAGATGAAGTAGTAACCTTACC 4084
Oy 3616 -----caatacaaaacggttagccggtacagccacagactggggcaagatcaaccatcga 3671
Db 4085 GATTTTAAGCAGACACAGGGTCTGGGCTTAGTCAAGAACTTTGACACGAGCAATACTCCA 4144
Oy 3672 agtalcgcctcagagatggaatccacatagaggggcagatactgtttccaactaaatcccg 3731
Db 4145 ACTTACACACTACATTAAGATACAGTACAGAAAGCTGTGATTAAGCGCTCCCAAGTCTTCAG 4204
Oy 3732 acagtgaatgaatgatactatctgcggtttcaaggactagtgcttctcglttcagca 3791
Db 4205 ATAGTGATGTGAGTGTGATGTCGCGCATTTCCAGAGCCACAGTACTCCAGCCTCAGCA 4264
Oy 3792 gcaagaagctacatgtctgcgtccaatcagagcgccgagaggaagaacaggaataatcgttct 3851
Db 4265 GGCAGAGCTTTATGTAGAGCACTGTGAGCGCCCCAGGG-----GTAGGATCAGTTTCAT 4318
Oy 3852 ttaacatccaaaatgcaaaaacagacagatggcggtgtcgggggaaagaaactgacaaaagca 3911
Db 4319 TTACCCCCCAAAATGCAAGGACAGAGCGATGGGACTTCAGGAAGAGCATCATCAAGAGCA 4378
Oy 3912 ccagcatcaatgagagacatgtctcaactggagaagaatgaacgcgcagccagtcgcgaactg 3971
Db 4379 CCAGTGTAGTGTGAGAGATATATACACTGGAACGTATGTAGCGGTAGCCAGTCCGAGACGG 4438
Oy 3972 cagtgaggcccttggttaccagtggaagaaggcgagctagatctggggccaaaatgg 4031
Db 4439 CCGTAGGTACCGCTGGAGCGGAGGAAAGAAAGATCCAGCCTGAGCGCCAAAGTGG 4498
Oy 4032 tagcatgtgtgctcctcaagaaagtcgacgtgtcctcaactcagccaaacggaag 4091

```

```

Db 4499 TAGCCATTG-----TGCTCGAAGAAGCAGAGCACGTCACAGCTTCAGCCAGACAGT 4552
Oy 4092 gaggaggttaaaaagcttaaggagcactgtctcagagaagcaaggagaccgggttagcagtg 4151
Db 4553 CGGGCCACAGAAAGTTGAAAGCACCATCCAGAGAGTACGAAACAGGAATGGACACTG 4612
Oy 4152 agatgagaaacttgatlgaccgccagccagccgggaatccacagatlgagcaatgaca 4211
Db 4613 AAATGCGGAA---GATGCGAGACAGCCGAGCGGGAGTGTCCAGAGATGGCAGATCAACA 4669
Oy 4212 gctatagctcggaagaaatctgacttcccttgggtgcgcttgcgctctgacagcagt 4271
Db 4670 GTTATAGCTCGGAAGGAAACTTGATTTCTTGAGTTGAGTAGGACCCGACAGTACGT 4729
Oy 4272 tcaagtattccttgaatgctgcgtggccctgtcagcttagtgaggagccagacccttgcta 4331
Db 4730 TCAGTGAATTCCTTGATGGGggttgggACACGCGAGCTGTTGGCGGTCAACGCTGCCA 4789
Oy 4332 ctccctgcaatgggtgacatcgaatgagtgaaatggaataaaagggaacgttgagtg 4391
Db 4790 CCGCGCCATGGGCGATATCCAAATCGGGATGGAGATGAAGAGGTCAATGGAGGTTG 4849
Oy 4392 aaatcatccgggcgcgcgccttggtaaaccaggttccaaagacatgcccagccgt 4451
Db 4850 AGCTTATCAGAGCCCGAGACCTTACACAAAACCTGTTCCAAATCTAACACCCCTCCT 4909
Oy 4452 atgtcaagtgatctgttagaacaagagtgatgcatagccaaagaataaaacaaagtg 4511
Db 4910 ATGTGAAATATATCTTTTGAAATGAGACCTGTATGGCAAAAAGAAAGCAAGAAATG 4969
Oy 4512 cgagaagaacccctggagccctgtaccagagctcttgccttcgagagagccccag 4571
Db 4970 CACGGAAACTCTCGATCTCTTGTATCAGCAGTCCCGGTTTGTGAGAAAGTCCACAG 5029
Oy 4572 ggaagggttlaacagatcgttcttgggagatagtgctglatgualacaaatccctta 4631
Db 5030 GTAAGTCTTCAGGTGATGTTGCTGGGTGATGGAAGATGACACCAAAATGCTTTA 5089
Oy 4632 tggagatggcccaataactcttagatgaatggaactgaactatccaaatggtatgagtg 4691
Db 5090 TGGGTGTGCTCAAACTTGTGGAAGAACTGTATCTATCCAGCATGTGATGTGATGT 5149
Oy 4692 tcaaatcttccctctctcctcctagtagatccaactgcgcaccccttgaagaagag 4751
Db 5150 ATAAATTTGTTCCCTCGCTCTCACTGTGATCCCACTCGCTCCCTACCCGCGGG 5209
Oy 4752 ctcccaalcgctctctggaagtlctacacggaccttactctcgtlctcatalagcaa 4807
Db 5210 CTTCCTCAATCTCTGGAAGTTTCGTCGGGCTCCTCCATCCGGTCAATAGTAA 5265

```

Search completed: November 21, 2001, 18:56:01
Job time: 10687 sec


```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Db 600 AGTCGTACATAGTGGCTAGCAAGG 627

RESULT 2
AA476826 590 bp mRNA EST 08-AUG-1997
LOCUS zW9501.r1 Soares total fetus ND2HF8.9w Homo sapiens cDNA clone
DEFINITION IMAGE:764752.5', similar to WP:TI0A3.1 CE04931 C2 DOMAIN OF PROTEIN
KINASE C ;, mRNA sequence.

ACCESSION AA476826 GI:2205037
VERSION AA476826
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 590)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
Schellenberg, K., Steptoe, M., Tan, F., Theisling, B., White, Y., Wyllie,
, T., Waterston, R. and Wilson, R.
Washu-Merck EST Project 1997
UNPUBLISHED (1997)
CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Seq primer: -28m13 rev2 EF from Amersham
High quality sequence stop: 495.
Location/Qualifiers

FEATURES
source
1..590
/organism="Homo sapiens"
/db_xref="GDB:5982370"
/db_xref="taxon:9606"
/clone="IMAGE:764752"
/clone_1lb="Soares_Total_fetus_ND2HF8.9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCCCTTAATTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 188 a 135 c 129 g 138 t

ORIGIN

Query Match 9.08; Score 446.4; DB 8; Length 590;
Best Local Similarity 89.48; Fred. No. 1.6e-114;
Matches 504; Conservative 0; Mismatches 56; Indels 4; Gaps 2;

Db 184 ACGGAGTCTCATAGCCAAAAGAAAACAAAGTGGCAAGAAAAACGTGACCCCTTT 243

QY 4536 accagagctcttgccttccttcgagagagagcccccagggaggggtgtatgagatcattgtc 4595
|||||

Db 244 ACCGAGAGCTATTATCTTTTGGAGAGAGTCCACAGGAAAAAGTTTTCAGATCATCGTCT 303
|||||

QY 4596 gggagagatlatgctcgtatgatgatacaaatcccttatggaggtggccagatcttag 4655
|||||

Db 304 GGGGAGATTATGGCCCGCATGATCACAATCTTTTAIOGGAGGTGCCAGATCTTTAG 363
|||||

QY 4656 atgaactggaactatccaacatggtatgttgatggttcaaacctctccctccctcc 4715
|||||

Db 364 ATGAAGTAGAGCTATCAATATATGATGATCGATGATGATCAACTTTTCCACCTTCCTCC 423
|||||

QY 4716 tagtagatccaactcggacactcttcgacaagaagagcttcccaatcgtctctggaagt 4775
|||||

Db 424 TAGTAGATCCAACTTGGCTTCCTCGACAGAAAGAGCTTCCCAATCATCTTGAAAAATT 483
|||||

QY 4776 ctaccgagctcttactctcgttcacagaaact---ataaactgtgtcacacaacc 4832
|||||

Db 484 CAACGTGACCTTTCTTACTCTTCGTTACATGACGCTGTATAAAAAATTGTTGTACACGACAC 543
|||||

QY 4833 agcgatacaaaaacccaagaagaaa 4856
|||||

Db 544 AGCGTTACAAAAAATAAAAAA 567

RESULT 3
AA602733/c 523 bp mRNA EST 08-OCT-1997
LOCUS nP1704.s1 NCI-CGAP_B-3 Homo sapiens cDNA clone IMAGE:1116630 3'
DEFINITION similar to WP:TI0A3.1 CE04931 C2 DOMAIN OF PROTEIN KINASE C ;, mRNA
sequence.

ACCESSION AA602733 GI:2436667
VERSION AA602733
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 523)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
UNPUBLISHED (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1223 Std Error: 0.00
Seq primer: -40m13 fwd. EF from Amersham
High quality sequence stop: 449.
Location/Qualifiers

FEATURES
source
1..523
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1116630"
/clone_1lb="NCI-CGAP_B-3"
/sex="female"
/tissue_type="breast tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: breast. Vector: Bluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Ductal breast tumor. 5' adaptor sequence: 5'
GAATTCGCCAGAG 3' 3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTTCTTTT 3' Average insert size: 0.9 kb. "

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-renal.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://lml.lnl.gov
Plate: LML10432 row: 1 column: 21
High quality sequence stop: 583.
Location/Qualifiers

FEATURES

source

1..588
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4525916"
/clone.lib="NIH_MGC_93"
/tissue.type="transitional cell papilloma, cell line"
/lab.host="DH10B (phage-resistant)"
/note="Organ: bladder; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 180 a 135 c 127 g 146 t
ORIGIN

Query Match 8.1%; Score 403.6; DB 153; Length 588;
Best Local Similarity 84.7%; Pred. No. 1.9e-102;
Matches 500; Conservative 0; Mismatches 84; Indels 6; Gaps 4;

4334 atcaccgagcgagcgcttggttaaacaggttccagacatgcagaccgtat 4453

1 ATCATCCGGCCCGCCGTCGTC-TGTGTAAACACAGGTTCCACAGCACTGCCAGCAGCAGTAT 58

4454 gtcaagagtgatctgttagacaacagagctcatagccaacaaagaacacagctggcg 4513

59 GTAAAGAGTATCTATTAGATACGAGAGCTGCTATGACCAAAAAGAAACAAAGTGGCA 118

4514 agaaagaccctggagccctctgacagacagctctgtccctgagagagagcccccaggg 4573

119 AGAAAAAGCGTGAACCCCTTACAGCAGCTATATCTTTCGACAGAGTCCACAAGGA 178

4574 aggggtgtacagatcatgtctggggagattatgtcgtatgatcaaatcccttattg 4633

179 AAGTTTACAGATATGCTGCGGAGATTTAGCCGCATGATCAAAATCTTTATG 238

4634 ggaatggccagatactcttagatgaactgtaactcaacatgggtgattgattgttc 4693

239 GGAGTGGCCCAATCTTTAGATGACAGCTATGCCAATATGGTATGGATGGATGTTTC 298

4694 aaactcttcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 4753

299 AAACCTTTCCACCTTCCCTAGATGATCAACCTTGCCCTGTGCAAAAGAGCT 358

4754 tcccaatgctctctggaaggtctacacgacctcttactctcgttcatagcaac--tat 4811

359 TCCCAATATCTCTGGAAGTTCACCTGACCTCTTACTCTCGTTCAATAGTAGCTGTA 418

4812 aaactgtctacacaacagcgatatacaaaacagaagaagaacagcagc-tggaag 4870

419 AAAAATTTGTGCACGACACGAGCTTACAAAACAAAAAATATCCCGGGTTGCAAA 478

RESULT 6
AA772858/c 474 bp mRNA EST 29-JAN-1998

LOCUS
DEFINITION
ab70a10.r1 Stratiagene fetal retina 937202 Homo sapiens cDNA clone

IMAGE:852282.5' similar to TR:Q92511 Q92511 MYELOBLAST KIAA0237. ;,
mRNA sequence.

ACCESSION
VERSION
AA772858
AA772858.1 GI:2625700

KEYWORDS
EST.
SOURCE
human.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
1 (bases 1 to 474)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,

Kilman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
J., Moore,B., Scheinberg,R., Steptoe,M., Tan,F., Theising,B.,

White,Y., Wylie,T., Waterston,R. and Wilson,R.
Washu-NCI human EST Project

TITLE
JOURNAL
Unpublished (1997)

COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.

Possible reversed clone: similarity on wrong strand
Seq primer: -26ml3 rev1 EF from Amersham

High quality sequence stop: 458.
Location/Qualifiers

1..474
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:852282"
/clone.lib="Stratiagene fetal retina 937202"
/sex="mixed"
/lab.host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site:1: EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Pooled retinal tissue. Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCAGAG 3' ~3' adaptor sequence: 5' CTGCACTTTTCTTTTCTTTT 3'"
BASE COUNT 111 a 107 c 110 g 146 t
ORIGIN

Query Match 7.9%; Score 395.6; DB 11; Length 474;
Best Local Similarity 89.7%; Pred. No. 3.2e-100;
Matches 425; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

4319 cagacccctgctactcctcgaatggtgacatcaggtgggaatgatgataaagaagg 4378

474 CAGACTGTGGCAACACCTGCATGGGTGACATTCAGGTAGSAAATGATGACAAAAAGGA 415

4379 caagctggagtagaatacatcctcggcgcgcccttggtgtaaaacaggtctccaagca 4438

414 CAGCTGGAGGTAGAAATCATGCCCCCGCTTGTGTTGTTAAACACGATTCACAAAGACA 355

4439 ctggccacaccgctgctgcaaggtgtatctgtttagaacaacggaatcgtgataagca 4498

354 CTGCACACCGCTGCTGTAAGGTATCTATTAGATTAAGGAGTGTGATGCAATAAAG 295

4499 aaaaacaggttgccgaagaagaccctggagccctgtacacgagcctctgtcctcgag 4558

294 AAAACAAAGTGGCAAGAAAAACCTGTGAACCCCTTACACGACGCTATTTCCTTGA 235

4559 gagagccccaaggagaggtgttacaagatcatgtctgggagattatggtcgtatgag 4618

234 GAGAGTCCACAGAGAAAGTTTACAGATCATGCTGCGGAGGAGATTATGCGCCATGAT 175

```

QY 4619 cacaatccttagtgaggagtgccacagatactcttagatgaactggaactatccacatg 4678
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 174 CACAAATCTTTATGAGAGTGGCCCAATACTTTAGATGAATGAGACTATCCAAATATG 115
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 4679 gtgatggatggttcaaaactcttccttccttccttccttagatgaatccacactcggaact 4738
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 114 GTGATGGATGATGTCGAACACTTTCCACACTTCCTCCCTAGTAGATCAACCTTGCCCTT 55
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 4739 ctgcaagaagaagacttcccaatcgctctctggaagttctaccggaactctctaac 4792
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 54 CTGCAAGAAGAGCTCCCAATCTCTCTGGAAGTTCACTGACCTCTTAC 1
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 7
LOCUS BE750318 571 bp mRNA EST 15-SEP-2000
DEFINITION 201646 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BE750318
VERSION BE750318.1 GI:10164310
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 571)
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
,W.W. and Keefe,J.W.
Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
Unpublished (2000)
JOURNAL Contact: Smith TPL
COMMENT USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGACG
Plate: 111 row: N column: 17
Seq primer: ATTTAGTGACACTATAG.
FEATURES
source
1. 571
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_id="MARC 4BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site 1: XbaI; Site 2: XhoI;
Library made from pooled tissue from day 20 and day 40
embryos."
location/Qualifiers
BASE COUNT 167 a 139 c 157 g 108 t
ORIGIN
Query Match 7.9%; Score 391.4; DB 139; Length 571;
Best Local Similarity 86.7%; Pred. No. 5.4e-99;
Matches 431; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
QY 3836 aggaataatcagtgctttacataaataagacaagacagatggcggtgcggggaag 3895
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 74 AAGAAATCCGTGCTTACTTCAATGCAAGCAGACAAATGGGCACTCCAGGGAAG 133
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3896 aacttgaccacaagaccagcatcagtgagacatggtctcactgagagaagaatgaagcgc 3955
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 134 AACATGACCAAGACCGACGATCAGCGAGACATGCTGCTGAGAGAAATGATGGC 193
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3956 agccagtcgcgactgacgtggtggccctgtgtaccatgtgcaagaagcgcgatctagc 4015
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

```

Db 194 AGCCAGTCGACACACCGAGTGGTGCTTGGGACACAGCACAGAAAGGACGCTTAGC 253
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 4016 attggggccaaatgtagctatgtgtgtctctcaggaataagtcgagtcctcaaa 4075
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 254 ATGGGGGCCAAATATGGATATCGTTGGTCTTCCGGAAGACCGCAGTGCCTCAG 313
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 4076 ctgagccaacccgaaagagaggttaaaaagcttaaggagcctgtctgagaagcagag 4135
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 314 CTCAGCCCAAGCAGAAACGAGAGGTAAGAAAGCTACGGAGCACTGTCCAGAGAAATACAAA 373
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 4136 accggctagcagtgagatgaagaactggaatgaccccgagcgaacgggaatccaca 4195
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 374 ACAGGCTTGCGTGTGGAATATGAGAGAACTGCATGACTGCAGACAGGACCGCAATCTTAC 433
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 4196 gatggacgataagacataagctcggaaagaaatcgtatcttccttgagtcgcgctg 4255
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 434 GACGGCAGCATGGAACACTACACTCAGAGGAAGAAATCGATCTCCCTGTCGCGCTTG 493
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 4256 gctctgacagcgacgtcagatgattctctgagtgcttgagcctgtcgaactagtggaa 4315
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 494 GCCTCGGATAGCCCAATTCAGATTTCTGGATGGCCTTGCCCTGCCACGACTGATGGGG 553
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 4316 cgccagaccctggctac 4332
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 554 CGCCAAACTCTGGCCAC 570
      ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 8
LOCUS BF791480 960 bp mRNA EST 12-JAN-2001
DEFINITION 602251523F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4344005 5',
mRNA sequence.
ACCESSION BF791480
VERSION BF791480.1 GI:12096534
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 960)
NIH-MGC http://imgc.ncl.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: L1AM961 row: n column: 06
High quality sequence stop: 445.
FEATURES
source
1. 960
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="IMAGE:4344005"
/clone_1ib="NIH_MGC_84"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: adrenal gland; Vector: PCMV-SPORT6; Site 1:
NotI; Site 2: SalI; Cloned unidirectionally; oligo-dr
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 344 a 215 c 239 g 162 t
ORIGIN
Query Match 7.4%; Score 366.6; DB 169; Length 960;

```

| SOURCE | ORGANISM |
|--------|---|
| human | Homo sapiens |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| | Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. |

FEATURES
source

Location/Qualifiers
1. .428
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3086882"
/clone_lib="NCI_CGAP_Sub8"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; NCI_CGAP_Sub8
is a subtracted library derived from NCI_CGAP_Sub5. The
NCI_CGAP_Sub8 library had 2.5 million recombinants. A
single-stranded DNA preparation of NCI_CGAP_Sub5 was used
as a tracer in a subtractive hybridization with a driver
clone comprising a pool of clones from NCI_CGAP_Sub5 (IMAGE
clone ids 2732833-2737415, 3068040-3069191; 25% of the
driver population), NCI_CGAP_Sub6 (pool AIF-AU, IMAGE ids
2728969-2733190; 25% of the driver population), and
NCI_CGAP_Sub7 (IMAGE ids 3069192-3072238, 3081664-3084550
; 25% of the driver population). Subtraction was
performed as previously described [Bonaldi, Lennon &
Soares (1996): Normalization and Subtraction: Two
Approaches To Facilitate Gene Discovery. Genome Research
6, 791-806.
TAG_LIB="NCI_CGAP_Lus
TAG_TISSUE="lung
TAG_SEQ="CAAC"

BASE COUNT 105 a 82 c 97 g 144 t

ORIGIN

Query Match 6.8%; Score 336.8; DB 149; Length 428;
Best Local Similarity 88.3%; Pred. No. 1.3e-83;
Matches 378; Conservative 0; Mismatches 47; Indels 3; Gaps 1;

QY 4431 ccaagacaccccgacacgcgtatgcaaggtgcatctgttacaacacggagctgcatag 4490
|||||
Db 428 CCAAGACGCTCGCACACCGTATGTAAGTATCTATCTATGATTAACGAGCTCGCATAG 369
|||||
QY 4491 ccaaaaagaacacaaaggtgagcaagaacccctgagccctgtacacagcagcttgt 4550
|||||
Db 368 CCAAAAAGAAACAAACAAAGTGGCAAGAAAACCGCTGGAACCCCTTACACAGCACTATAT 309
|||||
QY 4551 ccttcggaggaagcccccgaaggagggtgtacacatcatgtctgggagatagtc 4610
|||||
Db 308 CTTTCGAGAGAGTCCACAGGAAAAGTTTACAGATCATCTCTGGGAGATTTATGGCC 249
|||||
QY 4611 gtatgcatcacaatccttattgaggtgagccagatatctctagaactggaactat 4670
|||||
Db 248 GCATGATCACAAATCTTTATGGAGTGGCCACATATCTTTAGATGACTAGACTAT 189
|||||
QY 4671 ccaacatggtatgattgattgattcacaactctccctcctcctcctcctcctcctcct 4730
|||||
Db 188 CCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 129
|||||
QY 4731 cggacacctctgacaagaagagcttcccaatgctctctggaaggtcttaccggaacctct 4790
|||||
Db 128 TGGCTCTCTCTACAGAGAGCTTCCATCATCTCTGGAAGTTTCAATGTGACCTCTCTT 69
|||||
QY 4791 actctgcttcatagcaact---ataaacgtgttcacacacacacgagatacaaac 4847
|||||
Db 68 ACTCTCTCTCTATAGAGCTGTAAATAATGTGTTCACAGCAACCGCTTACAAAAA 9
|||||
QY 4848 agaagaaa 4855
|||||
Db 8 AAAAAAA 1

RESULT 14
BE980779 456 bp mRNA EST 05-OCT-2000
LOCUS BE980779
DEFINITION UI-M-BG2-bcn-a-01-0-UI.s1 NIH_BMAP_MSC_S1 Mus musculus cDNA clone

ACCESSION
BE980779
VERSION
BE980779.1 GI:10649193
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 456)
Bonaldi,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE
97044477
COMMENT
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@mail.nih.gov
Oligo-dt track not found. Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone Distribution: Researchers may obtain BMAP cDNA
clones from RESEARCH GENERICS. It should be noted that Bento Soares
is generating a small number of additional specialized
non-redundant arrays of BMAP cDNAs whose availability will be
considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA=No.

FEATURES
source

Location/Qualifiers
1. .456
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BG2-bcn-a-01-0-UI"
/clone_lib="NIH_BMAP_MSC_S1"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The
NIH_BMAP_MSC_S1 library is a subtracted library derived
from NIH_BMAP_MSC_N. NIH_BMAP_MSC_N was made from mouse spinal
cord tissue. For a detailed description of the library
from which this clone was derived, please visit our web
site at brainest.eng.uiowa.edu.
TAG_SEQ=None found"

BASE COUNT 129 a 116 c 97 g 114 t

ORIGIN

Query Match 6.7%; Score 332.4; DB 142; Length 456;
Best Local Similarity 98.2%; Pred. No. 2.3e-82;
Matches 336; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4639 ggcccaatactcttgaatgactggaactatccaactggtgattgattgattcaact 4698
|||||
Db 2 GGCGCGTACTCTTAAATGAATGGAACATCAACATGAGATTGATGATGATCAACT 61
|||||
QY 4699 ctccctccttccctccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 4758
|||||
Db 62 CTTCCTCCT 121
|||||
QY 4759 atcgctcttggaagttctaccggaaccttctactctcggttcatagaactataaactg 4818
|||||
Db 122 ATCGTCTCTGGAAGTTCTACCGGACCTTCTACTCTGTTCAATGCAACTATAAAC 181
|||||
QY 4819 ttgtcacacaacacagagatacaaaaacagaaagaaacagacaggttgaagcccttgt 4878
|||||
Db 182 TTGTCAACAACACGAGATCAAAACAGAGAAAACGACAGGTGGAAGCCCTGTGT 241
|||||
QY 4879 aacactgcatgattgattgtctacagaccacgcttaggagatccaaagcagctct 4938
|||||

Db 242 AACACTGACATGTTGATGTGTCTACAGAGCCACGCTAGGATACCAAGCAGTCT 301

QY 4939 ggtgttcagaggaagtcgtacacattgtgccttagcaagg 4980
 |||||||

Db 302 GTGTCTCAGAGGAAGTCGTACACATTGTCCCTAGCAAGG 343

RESULT 15
 BE955159

LOCUS BE955159 463 bp mRNA EST 04-OCT-2000

DEFINITION UI-M-BH4-bbf-b-06-0-UI.s1 NIH_BMAP_M_S5 Mus musculus cDNA clone

ACCESSION UI-M-BH4-bbf-b-06-0-UI 3', mRNA sequence.
 BE955159

VERSION BE955159.1 GI:10598417

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 463)

AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

COMMENT Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: messt@mail.nih.gov
 Oligo-dT track not found, Not 1 site shown in beginning of sequence is likely internal to the message. cDNA library preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
 Seq primer: M13 forward
 POLA=NO.

FEATURES
 source Location/Qualifiers
 1..463
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-BH4-bbf-b-06-0-UI"
 /clone_lib="NIH_BMAP_M_S5"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="vector: p77T3D-Pac (Pharmacia) with a modified polylinker Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M_S5 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. For a detailed description of the library from which this clone was derived, please visit our web site at brainest.eng.uiowa.edu.
 TAG_SEQ=None found"

BASE COUNT 131 a 118 c 99 g 115 t

ORIGIN

Db 2 GGCCGCTTACTCTTAGATGAACCTGGAACTATCCAACTGGTGGATTGGATGGTTCAACT 61

QY 4699 ctccctctctccctccctagtagatccaaacctcgacacctgacaagaagacctcca 4758
 |||||||

Db 62 CTCCCTCATTCCTCCCTAGTAGATCCAACTTGGACACTGTGACAAGAGAGCTTCCCA 121

QY 4759 atggtctctggaagttctaccggacctcttacctctggttataatgcaactataaactg 4818
 |||||||

Db 122 ATCGTCTCTGAAAAGTTCTACCGGACCTTCTACCTGTTCTATGCAACTATTAACCTG 181

QY 4819 ttgtacacacacacagcgatcaaaaacccaagaagaacacaggtggaagccctgtg 4878
 |||||||

Db 182 TTGTACACAACACGCGCATACAAAACCAAGAGAAAACGACAGGTGGAAGCCCTGCT 241

QY 4879 aacactgacatctgtagtctgtctacagagccacgtctaggatataccaagcagtcct 4938
 |||||||

Db 242 AACACTGATCTTGATGTGTCTACAGAGCCACGCTAGGATACCAAGCAGTCT 301

QY 4939 ggtgttcagaggaagtcgtacacattgtgccttagcaagg 4980
 |||||||

Db 302 GTGTCTCAGAGGAAGTCGTACACATTGTGCCCTAGCAAGG 343

Search completed: November 21, 2001, 20:05:52
 Job time: 14878 sec

Query Match 6.7%; Score 332.4; DB 142; Length 463;
 Best Local Similarity 98.2%; Pred.No. 2.3e-82;
 Matches 336; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4639 ggccaatactcttagtgaactgaactatccacaatgtagtgaatggtcaact 4698
 |||||

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 21, 2001, 20:56:50 ; Search time 143.6 Seconds
(without alignments)
7854.181 Million cell updates/sec

Title: US-09-617-099B-2
Perfect score: 4980
Sequence: 1 gcttcctaggtgtgttcg.....acattgtgcctagaagaag 4980

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgnl_7/prodata/1/lna/5A_COMB.seq: *
2: /cgnl_7/prodata/1/lna/5B_COMB.seq: *
3: /cgnl_7/prodata/1/lna/5A_COMB.seq: *
4: /cgnl_7/prodata/1/lna/5B_COMB.seq: *
5: /cgnl_7/prodata/1/lna/PCTUS_COMB.seq: *
6: /cgnl_7/prodata/1/lna/Dackfilesl.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------------------|--------------------|
| 1 | 67.2 | 1.3 | 7218 | 1 US-08-232-463-14 | Sequence 14, Appl |
| 2 | 47.6 | 1.0 | 2244 | 3 US-09-094-714A-48 | Sequence 48, Appl |
| 3 | 47.6 | 1.0 | 2245 | 4 US-09-225-749-24 | Sequence 24, Appl |
| 4 | 46 | 0.9 | 533 | 6 5482709-5 | Patent No. 5482709 |
| 5 | 45 | 0.9 | 3489 | 2 US-08-728-323A-1 | Sequence 1, Appl |
| 6 | 45 | 0.9 | 32207 | 2 US-08-770-379-20 | Sequence 20, Appl |
| 7 | 45 | 0.9 | 32207 | 4 US-08-757-669A-20 | Sequence 20, Appl |
| 8 | 43.4 | 0.9 | 397 | 3 US-09-253-691-3 | Sequence 3, Appl |
| 9 | 42 | 0.8 | 203 | 4 US-09-043-301-7 | Sequence 7, Appl |
| 10 | 41.4 | 0.8 | 51259 | 3 US-08-781-891-209 | Sequence 209, App |
| 11 | 40.8 | 0.8 | 477 | 4 US-09-135-994-1 | Sequence 1, Appl |
| 12 | 40.8 | 0.8 | 543 | 6 5273901-6 | Patent No. 5273901 |
| 13 | 40.8 | 0.8 | 3891 | 4 US-09-036-315-1 | Sequence 1, Appl |
| 14 | 40.6 | 0.8 | 16442 | 3 US-08-781-891-208 | Sequence 208, App |
| 15 | 39.6 | 0.8 | 154 | 1 US-08-469-802B-6 | Sequence 6, Appl |
| 16 | 39.6 | 0.8 | 154 | 2 US-08-267-803B-6 | Sequence 6, Appl |
| 17 | 39.6 | 0.8 | 165 | 4 US-09-043-303-17 | Sequence 17, Appl |
| 18 | 39.6 | 0.8 | 168 | 1 US-08-469-802B-4 | Sequence 4, Appl |
| 19 | 39.6 | 0.8 | 168 | 2 US-08-267-803B-4 | Sequence 4, Appl |
| 20 | 39.6 | 0.8 | 171 | 1 US-08-469-802B-5 | Sequence 5, Appl |
| 21 | 39.6 | 0.8 | 171 | 2 US-08-267-803B-5 | Sequence 5, Appl |
| 22 | 39.6 | 0.8 | 195 | 1 US-08-469-802B-2 | Sequence 2, Appl |
| 23 | 39.6 | 0.8 | 195 | 2 US-08-267-803B-2 | Sequence 2, Appl |
| 24 | 39.6 | 0.8 | 234 | 1 US-08-469-802B-3 | Sequence 3, Appl |
| 25 | 39.6 | 0.8 | 234 | 2 US-08-267-803B-3 | Sequence 3, Appl |
| 26 | 39.6 | 0.8 | 3376 | 1 US-08-320-559-29 | Sequence 29, Appl |
| 27 | 39.6 | 0.8 | 3376 | 3 US-08-545-860D-29 | Sequence 29, Appl |

| | | | | | |
|----|------|-----|--------|---------------------|-------------------|
| 28 | 39.6 | 0.8 | 3376 | 5 PCT-US94-04496-29 | Sequence 29, Appl |
| 29 | 38.4 | 0.8 | 2301 | 1 US-08-306-691B-23 | Sequence 23, Appl |
| 30 | 38.4 | 0.8 | 2301 | 5 PCT-US93-06251-78 | Sequence 78, Appl |
| 31 | 37.8 | 0.8 | 289 | 4 US-09-007-005-17 | Sequence 17, Appl |
| 32 | 37.8 | 0.8 | 289 | 4 US-09-244-796-17 | Sequence 17, Appl |
| 33 | 37.8 | 0.8 | 1441 | 1 US-08-136-277-18 | Sequence 18, Appl |
| 34 | 37.8 | 0.8 | 1441 | 2 US-08-479-403-18 | Sequence 18, Appl |
| 35 | 37.8 | 0.8 | 1441 | 3 US-08-835-734-18 | Sequence 18, Appl |
| 36 | 37.4 | 0.8 | 1994 | 2 US-08-933-750C-58 | Sequence 58, Appl |
| 37 | 37.4 | 0.8 | 1994 | 3 US-09-234-613-58 | Sequence 58, Appl |
| 38 | 37 | 0.7 | 441529 | 4 US-09-103-840A-1 | Sequence 1, Appl |
| 39 | 36.6 | 0.7 | 921 | 2 US-08-795-475-2 | Sequence 2, Appl |
| 40 | 36.6 | 0.7 | 1308 | 2 US-08-795-475-4 | Sequence 4, Appl |
| 41 | 36.2 | 0.7 | 5285 | 2 US-08-609-049A-29 | Sequence 29, Appl |
| 42 | 36.2 | 0.7 | 5285 | 4 US-09-170-996-29 | Sequence 29, Appl |
| 43 | 36 | 0.7 | 3337 | 1 US-08-072-610-1 | Sequence 1, Appl |
| 44 | 36 | 0.7 | 3337 | 2 US-08-719-822B-1 | Sequence 1, Appl |
| 45 | 36 | 0.7 | 3337 | 4 US-09-092-458-1 | Sequence 1, Appl |

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9pt-F15
US-08-232-463-14

Query Match 1.3%; Score 67.2; DB 1; Length 7218;
Best Local Similarity 6.4%; Pred. No. 1.1e-09;
Matches 27; Conservative 232; Mismatches 165; Indels 0; Gaps 0;

1648 taatccaagaagaagtagtaaaatgctccaggttctactgagcaactggaggaagagct 1707
1437 TACRR 1378
1708 ggcctccacccctagatatacaagctgcatgagtgagcgaagcgaagtgtag 1767
1377 RRR 1318
1768 tgaagaagggagacatcaaaagggaagaaagaaactagtgagcgaaggtttgtcgaa 1827
1317 RRR 1258
1828 ttctaaccggctctgagagacaaagaaagagatgactctggtgagccactcttga 1887
1257 RRR 1198
1888 agaagattggaatgctgagccctcagattaagagacttgggtagatcctgtagtag 1947
1197 RRR 1138
1948 cacaaccccttaacgagagcagatagccatagtgataagcaacctgtgactgagcagccatc 2007
1137 RRR 1078
2008 caaagatggaagatgccttaatgctgatttatataaagcgttaaaagatggag 2067
1077 RRRRRRRRRRRATCCAGACTCCCTCAGCTCGACCTGACGCAAGCTCGAATTAATCTGTGAG 1018

2068 tglta 2071
1017 CGTA 1014

RESULT 2
US-09-094-714A-48
Sequence 48, Application US/09094714A
Patent No. 6117847
GENERAL INFORMATION:
APPLICANT: C. Frank Bennett, Nicholas M. Dean
TITLE OF INVENTION: OLIGONUCLEOTIDES FOR ENHANCED MODULATION OF
PROTEIN KINASE C EXPRESSION
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 6117847rls, LLP
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERECT 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/094,714A
FILING DATE: June 15, 1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/601,269
FILING DATE: 14-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/478,178
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/089,996
FILING DATE: 09-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/852,852

FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Paul K. Legard
REGISTRATION NUMBER: 38,534
REFERENCE/DOCKET NUMBER: ISIS-2943
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ. ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 2244
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-094-714A-48

Query Match 1.0%; Score 47.6; DB 3; Length 2244;
Best Local Similarity 48.5%; Pred. No. 0.00054;
Matches 131; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

2489 ataactatggttgacaaggttggtcaccagttgatagttacatttgggagcaag 2548
511 ATTACCTTAAGGCTGAGGTCTGATGAAAGCTCATGTCACGTACGAGATGCANAA 570
2549 gatctcccttcagggaagatgtaggagccaaagaaatcctatgttaagtattcctt 2608
571 AATCTAATCCCTATGAGATCCAAACGGCTTTCAGATCCTTATGTGAGTGAACCTTAT 630
2609 ccagatagaagtgtataaataaagaagaacaacaaacagtcgaagaaacttgaaccc 2668
631 CCTGATCCCAAGAAATGAAGACGACAAAACCAAAACCATCGCTCCACACTAATCTCG 690
2669 aaatggaaccgacttatttattctctcctcgcacgaagaatctcgtgaacgaatg 2728
691 CAGTGAATGAGTCTTACATTCAAATTTGAACCTTCAGACAAAGACGAGACATGTCT 750
2729 ctggaattacccttgggataagctaga 2758
751 GTAGAAATCTGGGACTGGGATGCAACACA 780

RESULT 3
US-09-225-749-24
Sequence 24, Application US/09225749
Patent No. 6300320
GENERAL INFORMATION:
APPLICANT: Dean, Nicholas M.
TITLE OF INVENTION: Modulation of c-jun using inhibitors of protein kinase C
FILE REFERENCE: ISIS313
CURRENT APPLICATION NUMBER: US/09/225,749
NUMBER OF SEQ. ID NOS: 24
SOFTWARE: PatentIn Ver. 2.0
SEQ. ID NO. 24
LENGTH: 2245
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (28)..(2046)
PUBLICATION INFORMATION:
JOURNAL: Nucleic Acids Res.
VOLUME: 18
ISSUE: 8
PAGES: 2183
DATE: 1990-04-25
DATABASE ACCESSION NUMBER: X52479/genbank
DATABASE ENTRY DATE: 1993-09-12
US-09-225-749-24

| | | | |
|----|------|--|------|
| QY | 349 | aatgataagagagcgaagtccaagaaga ttggaaagggaattcgcaacagcaagcaagaagacga | 408 |
| | | | |
| Dd | 2169 | GCAGCAGCGAGATGACACACGACGATGACACGACACGACGATGACGCA | 2228 |
| QY | 409 | gggtgatgcgccg | 421 |
| | | | |
| Dd | 2229 | GCAGGATGAGCAG | 2241 |

RESULT 6

```

US-08-770-379-20/c
Sequence 20, Application US/08770379
Patent No. 5849564
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ. ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

```

| | | | |
|----|-------|--|-------|
| Db | 19828 | GCAGCAGCAGGATGTAGCAGCAGCAGCAGGAGTATGACAGCAGCAGCA | 19769 |
| QY | 409 | gggtgatgccccg | 421 |
| | | | |
| Db | 19768 | GCAGGATGAGCAG | 19756 |

RESULT 7

```

US-08-757-669A-20/c
: Sequence 20, Application us/08757669A
: Patent No. 6183751
:
: GENERAL INFORMATION:
:
: APPLICANT: Chang, Yuan
:
: APPLICANT: Bohenzky, Roy A.
:
: APPLICANT: Russo, James J.
:
: APPLICANT: Edelman, Isidore S.
:
: APPLICANT: Moore, Patrick S.
:
: TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
:
: TITLE OF INVENTION: SEQUENCES AND USES THEREOF
:
: NUMBER OF SEQUENCES: 20
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: Cooper & Dunham LLP
:
: STREET: 1185 Avenue of the Americas
:
: CITY: New York
:
: STATE: New York
:
: COUNTRY: U.S.A.
:
: ZIP: 10036
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/08/757,669A
:
: FILING DATE:
:
: CLASSIFICATION: 424
:
: ATTORNEY/AGENT INFORMATION:
:
: NAME: White, John P.
:
: REGISTRATION NUMBER: 28,678
:
: REFERENCE/DOCKET NUMBER: 45185-F
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: (212) 278-0400
:
: TELEFAX: (212) 391-0525
:
: INFORMATION FOR SEQ. ID NO: 20:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 32207 base pairs
:
: TYPE: nucleic acid
:
: STRANDEDNESS: double
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: DNA (genomic)
:
: IS-08-757-669A-20

```

| Query Match | Similarity | 0.98 | Score 45 | DB 4 | Length 32207 |
|-------------|---|-------|-----------------|--------|--------------|
| Best Local | Similarity | 48.68 | Pred. NO. 0.027 | | |
| Matches 123 | Conservative | 0 | Mismatches 130 | Indels | Gaps |
| QY 169 | tcagaagaaagaaagagagaagacacagtcctcgctgcagaatcaagaatcaagaagaacacaaagc | 228 | | | |
| Db 20008 | TCAGACACACAGAGATGAGCGAGCAGAGATGAGCAGCAGCAGCAGATGAGCAGACGACCA | 19949 | | | |
| QY 229 | acaaagcaacagatgagttcccttattggaatcaatcgaactggttaataacgttctgca | 288 | | | |
| Db 19948 | TGAGCACACAGCAGATGACCGACGAGATGACAGCAGCAGATGAGCGAGGAGCACCA | 19889 | | | |
| QY 289 | gccccaagaaaaacaacccaatagaaagagccccaagaacaagctgcaccaacaattgca | 348 | | | |
| Db 19888 | GGATGAGCAGCAGCAGCAGCAGATGAGCAGGAGCAGCAGCAGATGAGCAGCAGGATGACCA | 19829 | | | |
| QY 349 | aatggtataagagacaagtccaagaagatgaggagaagaaatcgcaagcagcagaagccagaa | 408 | | | |
| Db 19828 | GCAGCAGCAGCAGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCGACCA | 19769 | | | |

[illegible]

RESULT 11
US-09-135-994-1
; Sequence 1, Application US/09135994A

```

: GENERAL INFORMATION:
: APPLICANT: Rannum et al.
: TITLE OF INVENTION: SCAT GENE AND METHODS OF USE
: FILE REFERENCE: University of Minnesota
: CURRENT APPLICATION NUMBER: US/09/135,994A
: CURRENT FILING DATE: 1998-08-18
: EARLIER APPLICATION NUMBER: 60/056,170
: EARLIER FILING DATE: 1997-08-19
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 1
: LENGTH: 477
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-135-994-1

```

| | | | | |
|-----------------------|--------------|------------------|----------------|-------------|
| Query Match | 0.88; | Score 40.8; | DB 4; | length 477; |
| Best Local Similarity | 55.78; | Pred. No. 0.021; | | |
| Matches 78; | Conservative | 0; | Mismatches 62; | Indels 0; |
| | | | Gaps | 0 |

[illegible]

QY 404 cagaaggtgatgtcccgac 423
||| | | ||| |
Db 274 cagcagcagcagccgccc 293

RESULT 12

Patent No. 5273901
APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
SPOROZOITE 21.5 KB ANTIGEN, AC-6B
NUMBER OF SEQUENCES: 11
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/581,693
FILING DATE: 12-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 215,162
FILING DATE: 05-JUL-1988
APPLICATION NUMBER: 746,520
FILING DATE: 19-JUN-1985
APPLICATION NUMBER: 627,811
FILING DATE: 05-JUL-1984
SEQ ID NO: 6
LENGTH: 543
5273901-6

| | | | | |
|-----------------------|--------------|------------------|-----------------|-------------|
| Query Match | 0.88; | Score 40.8; | DB 6; | length 543; |
| Best Local Similarity | 48.3%; | Pred. No. 0.023; | | |
| Matches 114; | Conservative | 0; | Mismatches 122; | Indels 0; |
| | | | Gaps | 0. |

[illegible]

RESULT 13
US-09-036-315-1
; Sequence 1, Application US/09036315
; Patent No. 6218523

```

?
? GENERAL INFORMATION:
? APPLICANT: French, Cynthia K.
? APPLICANT: Schneider, Patrick A.
? APPLICANT: Yamamoto, Karen K.
? TITLE OF INVENTION: Prostate Cancer-Specific Marker
? NUMBER OF SEQUENCES: 27
? CORRESPONDENCE ADDRESS:
?
```

ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California

; ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

;
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

; SOFTWARE:  FALCONER RELEASE #1.0
CURRENT APPLICATION DATA:

```

APPLICANT NUMBER: US/09/036,315

FILING DATE: 06-MAR-1998

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/041,811
FILING DATE: 15-MAY-1997

FILED DATE: 13-MAY-1997

APPLCATION NUMBER: US 60/041,246

FILING DATE: 07-MAR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Storella, John R.

REGISTRATION NUMBER: 32,944

REFERENCE/DOCKET NUMBER: 018002-00021005

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200

TELEPHONE: (415) 576-0300
TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 1

SEQUENCE CHARACTERISTICS:

LENGTH: 3891 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: CDNA
; TOPOLOGY: Linear

```

; MOLECULE TYPE: CDNA
FEATURE:

```

| NAME/KEY: | CDS |
|-----------|-----|
|-----------|-----|

LOCATION: 151..1425

OTHER INFORMATION: /product= "Repro-PC-1.0"

US-09-036-315-1

| | | | | |
|-----------------------|--------|-----------------|-------|--------------|
| Query Match | 0.88; | Score 40.8; | DB 4; | Length 3891; |
| Best Local Similarity | 54.7%; | Pred. No. 0.11; | | |

```

Matches      81: Conservative      0: Mismatches      67: Indels      0: Gaps      0:
Oy  4410 gccctgtgtgtaaaccaggttccaagacactgccaagcacgtaatgtaagytatctgt 4469
Db  1085 GACATCTGCTCAATAATCTGATGTGTCCGACTTTCAGATCCCTATGTCAAAAGTGAACCTGT 1144
Oy  4470 tagcaacagcggtctgtcatagcaaaagaagaaccaggtgtcgagaagaagccctggagc 4529
Db  1145 ACCATGCCCAAAAAGAGATCTCCCAAGAGAAGACTCATGTAAGAAGAACCCCAATG 1204
Oy  4530 cccgtaccagcagctctgtccctcga 4557
Db  1205 CAGTCTTCATGACGACTGTTTGTCTTTGA 1232

RESULT      14
US-08-781-891-208/c
: Sequence 208, Application US/08781891
: Patent No. 6090620
:
GENERAL INFORMATION:
: APPLICANT: Fu, Ying-Hui
: APPLICANT: Yu, Chang-Bn
: APPLICANT: Oshima, Junko
: APPLICANT: Mulligan, John T.
: APPLICANT: Schellenberg, Gerald D.
: TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
: TITLE OF INVENTION: WERNER'S SYNDROME
: NUMBER OF SEQUENCES: 209
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED AND BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: USA
: ZIP: 98104-7092
:
COMPUTER READABLE FORM:
:
MEDIUM TYPE: Floppy disk
:
COMPUTER: IBM PC compatible
:
OPERATING SYSTEM: PC-DOS/MS-DOS
:
SOFTWARE: Patentin Release #1.0, Version #1.30
:
CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/781,891
: FILING DATE: 27-DEC-1996
:
CLASSIFICATION: 800
:
ATTORNEY/AGENT INFORMATION:
: NAME: Mr. 6090620tenburg Ph.D., Carol
: REGISTRATION NUMBER: 39,317
: REFERENCE/DOCKET NUMBER: 240052.419
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 208:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 16442 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
US-08-781-891-208

Query Match      0.8%; Score 40.6; DB 3; Length 16442;
Best Local Similarity 47.8%; Pred. No. 0.37;
Matches 118; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Oy  163 ggaatcgctagaagaagaagaagagagagcagtcgctgtcctaagatcaagaagaaga 222
Db  16433 GGAGGACGACGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCA 16374
Oy  223 caaagcaacaacgcaacagtggttccctttagtgtgatacactgtaactgtaataacgt 282
Db  16373 GGAGCAGAGAGCAGCAGGAGGAGGAGCAGGACAGCAGGAGCAGCAGGAGGAGGAGCAGCA 16314
Oy  283 tctgcagccccaagaacaaacccaatgtgagaaggagccccaagacaagctgtaccaca 342

```

```

Db      161313 GAGGACAGGAGGAGGAGGAGCAGGAGAGGAGGAGGAGCAGGAGGAGGAGGAGGAGCAGGA 16254
QY      343 attgaatgtataagagcaagtccaagaatgtagaggaaattgcacgacagcagaaga 402
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      16253 GGAGCAGGAGGAGGAGGAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTA 16194
QY      403 gcgagaag 409
          | | | | |
          US-08-469-802B-6
RESULT 15
; Sequence 6, Application US/08469802B
; Patent No. 5741645
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Rannum, Laura P.W.
; APPLICANT: Chung, Ming-Yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene sequence for Spinocerebellar Ataxia
; Patent No. 5741645
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; CORRESPONDENCE ADDRESSES:
; ADDRESSSEE: Mueeling, Raasch, Gebhardt & Schwappach, P.A.
; STREET: 119 No. 5741645th Fourth Street, Suite 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,802B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mueeling, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 110,00030101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1225
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 154 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-469-802B-6
Query Match           0.8%; Score 39.6; DB 1; Length 154;
Best Local Similarity 57.1%; Pred.No.0.021;
Matches   72: Conservative    0; Mismatches   54; Indels     0; Gaps     0;
DB      284 ctgcagccccaacaacaaaccaatggaagagggcccagaacaagctgtaccacaaga 343
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
          5 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 64
QY      344 tttagaatgtataagagcgcaagtccaagaatgtagaggaaattgcacgacagcagaag 403
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      65 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 124
QY      404 cagaag 409
          | | | | |
          125 CAGCAG 130

```

Fri Nov 23 10:08:47 2001

Search completed: November 21, 2001, 23:59:49
Job time: 10979 sec

us-09-617-099b-2.rni

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2001, 16:03:34 ; Search time 29.74 Seconds
(without alignments)
4072.547 Million cell updates/sec

Title: US-09-617-099b-1

Perfect score: 8285

Sequence: 1 MSAPLGPGRPAVPAPASQP.....TRRASQSLESSTGSPYSRS 1590

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : PIR-68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|----------|---------------------|
| 1 | 4489.5 | 54.2 | 1553 | 2 T03301 | rab3 effector prot |
| 2 | 1002 | 12.1 | 833 | 2 T29736 | hypothetical prote |
| 3 | 500.5 | 6.0 | 743 | 2 T00634 | hypothetical prote |
| 4 | 475.5 | 5.7 | 1212 | 2 T00332 | hypothetical prote |
| 5 | 437 | 5.3 | 1325 | 2 T25753 | hypothetical prote |
| 6 | 358 | 4.3 | 547 | 2 T34318 | hypothetical prote |
| 7 | 338 | 4.1 | 1021 | 2 S44644 | hypothetical prote |
| 8 | 307 | 3.7 | 1898 | 1 A45973 | trichomyalin - hum |
| 9 | 278.5 | 3.4 | 1386 | 2 T49316 | proliferation relat |
| 10 | 272.5 | 3.3 | 5327 | 2 T13564 | microtubule-associ |
| 11 | 266 | 3.2 | 1560 | 2 T42727 | proliferation pote |
| 12 | 265 | 3.2 | 2526 | 2 T20531 | hypothetical prote |
| 13 | 264 | 3.2 | 681 | 2 JX0338 | rabphilin-3A - mou |
| 14 | 262.5 | 3.2 | 684 | 2 T58166 | rabphilin-3A - rat |
| 15 | 261.5 | 3.2 | 704 | 2 A48097 | rabphilin-3A - bov |
| 16 | 258.5 | 3.1 | 2843 | 1 RBH0AP | adenomatous polyo |
| 17 | 254.5 | 3.1 | 1845 | 2 T49505 | hypothetical prote |
| 18 | 253.5 | 3.1 | 1280 | 2 T00365 | hypothetical prote |
| 19 | 252 | 3.0 | 1791 | 2 T02345 | hypothetical prote |
| 20 | 246 | 3.0 | 1871 | 2 D96796 | probable heat shoc |
| 21 | 245 | 3.0 | 5170 | 2 T15348 | hypothetical prote |
| 22 | 243.5 | 2.9 | 1507 | 2 B47328 | natural killer cel |
| 23 | 242.5 | 2.9 | 2248 | 2 A35938 | profilagin - hum |
| 24 | 241 | 2.9 | 1549 | 1 A40691 | trichomyalin - she |
| 25 | 240 | 2.9 | 3147 | 2 T18674 | hypothetical prote |
| 26 | 238.5 | 2.9 | 1017 | 2 T15598 | hypothetical prote |
| 27 | 237 | 2.9 | 2722 | 2 T20532 | hypothetical prote |
| 28 | 237 | 2.9 | 2094 | 2 S33124 | tpi protein - huma |
| 29 | 236.5 | 2.9 | 1830 | 2 A37981 | microtubule-associ |

| | | | | | |
|----|-------|-----|------|----------|--------------------|
| 30 | 236 | 2.8 | 1825 | 2 S13507 | microtubule-associ |
| 31 | 234.5 | 2.8 | 1323 | 2 T00037 | hypothetical prote |
| 32 | 234.5 | 2.8 | 2938 | 2 T30249 | cell proliferation |
| 33 | 234 | 2.8 | 2207 | 2 T42759 | munc13-3 protein - |
| 34 | 233.5 | 2.8 | 1407 | 1 S28589 | trichomyalin - rab |
| 35 | 233.5 | 2.8 | 1558 | 2 T29253 | hypothetical prote |
| 36 | 231 | 2.8 | 1827 | 2 T16270 | hypothetical prote |
| 37 | 230.5 | 2.8 | 2453 | 2 S60254 | nuclear receptor c |
| 38 | 228 | 2.8 | 1217 | 2 S52714 | sericinb - silkwo |
| 39 | 227.5 | 2.7 | 2282 | 2 T42717 | DNA-binding protei |
| 40 | 227 | 2.7 | 1359 | 2 T34036 | hypothetical prote |
| 41 | 227 | 2.7 | 1695 | 2 T19823 | hypothetical prote |
| 42 | 225.5 | 2.7 | 1306 | 2 T13592 | hypothetical prote |
| 43 | 225.5 | 2.7 | 1829 | 2 T26135 | hypothetical prote |
| 44 | 225 | 2.7 | 1403 | 1 A47328 | natural killer cel |
| 45 | 225 | 2.7 | 1684 | 2 J00057 | gravin - human |

ALIGNMENTS

| | |
|---|---------------------------------|
| RESULT 1 | |
| T03301 | rab3 effector protein Rim - rat |
| C:Species: Rattus norvegicus (Norway rat) | |
| C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999 | |
| C:Accession: T03301 | |
| R:Wang, Y.; Okamoto, M.; Schmitz, F.; Hofmann, K.; Sudhof, T.C. | |
| Nature 388, 593-598, 1997 | |
| A>Title: Rim is a putative Rab3 effector in regulating synaptic-vesicle fusion. | |
| A:Reference number: Z14897, PMID:7394473 | |
| A:Accession: T03301 | |
| A>Status: preliminary; translated from GB/EMBL/DBJ | |
| A:Molecule type: mRNA | |
| A:Residues: 1-1553 <MAN> | |
| A:Cross-references: EMBL:AF007836, NID:q2317777; PIDN:AAB66703.1; PID:q2317778 | |
| A:Experimental source: tissue-type brain | |
| C:Genetics: | |
| A:Note: RIM | |
| C:Function: Rim protein is proposed as Rab3-dependent regulator of synaptic-vesicle | |
| A:Description: GTP binding; zinc finger | |
| C:Keywords: GTP binding; zinc finger | |
| Query Match 54.2%; Score 4489.5; DB 2; Length 1553; | |
| Best Local Similarity 57.2%; Pred. No. 2; 2e-214; | |
| Matches 952; Conservative 209; Mismatches 319; Indels 183; Gaps 37; | |
| QY 1 MSAPLGPGRPAVPAPASQPPEMPDLSTLTEREKIIIAVMDROKKEEKEBQSLKI 60 | |
| DB 1 MSSAVGPRGRPPPT---VPPMOBELDLSLTEREIIIAVMDROKKEEKEBQSLKI 56. | |
| QY 61 KEHKAQP-----TQWPFSSGITELVNNVLOPOOKOPNEKEPOT--KLHOQFEMYKE 110 | |
| DB 57 VVRDMKPAKCKPRMAESQPHQPRPLNIFRCVCVBRKRSSEGGGERDWRMLHQGESEYKE 116 | |
| QY 111 QVKKMGESQOOQ-EQGDAPTCGICHTKTRPADCGGHNCSQTKRFGKRCGRVSLRKNK 169 | |
| DB 117 QVKKIGEARKEGHEKDDAPTCGICHTKTRPADCGGHNCSQTKRFGKRCGRVSLRKNK 176 | |
| QY 170 ----VMMVNLCKRQOELLTKSGAMFYNSGSNTLQOPQKVPGRGLRN-----EAOE 218 | |
| DB 177 EDKVMVAVCNLCRKQOELLTKSGAMFYNSGSNTLQOPQKVPGRGLRN-----EAOE 218 | |
| QY 219 KKAHLADPOFOGAPGDLSPVAVEGRAHGLTRDQITKNGS-----GVKH 263 | |
| DB 231 KKAHLADPOFOGAPGDLSPVAVEGRAHGLTRDQITKNGS-----GVKH 263 | |
| QY 264 QTA-----SDMPDRKRKSPSVSRDQNRKREGESEKEDISQIV--PSDGTMPSPSPYADRR 317 | |
| DB 280 KQASRSRSEPPREKKAAGLS-EQNGKGQKSEKRVKPSVVOQEGT-----ADDERRK 333 | |
| QY 318 SQREPPQFYEEPGHILNVDSNNRGRHSKEYLVDDVDVSRDEYERQREEEYQARYRSDP 377 | |

Db 811 -WRTTQOOGSLIOMPMQMOOQHNTF-----DARWPKEDALSRMYATASRRRAOETA 861

QY 639 -----STTLEHSHS--DKHPVWOPSKDDRLIGRILLNKL 675

Db 862 LTSSMKISYGSRSYARRPIRPSYRNDEATNSMDRI-VARPIAENSRDYKCLITLTPSY 920

QY 676 KDGSVPBDSGAMLGKLVYGGKMTESGRLCATFKVKKGLADYGNLRPGDEYLEMNGRL 735

Db 921 KHNHNYND---LGVRVVGKKROMNGELSAVYSQLHSTANQTLGQIKIGDEYVEWNGIL 976

QY 736 LGATFFEEYVNIILEKRPQVEYLVSRIPIRPIRPNHQAQLESSSSSPESQKMDPS 795

Db 977 LRGKTFEEYERYL--NKSGLDEMT-----RTYKNPSSGYIDTLPLRN- 1019

QY 796 ISVTPSPSPGMLRDYPOF-----LSGOLSIKLMFDKVGHOIIVTI 835

Db 1020 -TMRDRLSPDVRPVYRMHNRINGINNNSVLHNHTLSDSSCHGHIQVLSGYDG-NSRLVAKI 1077

QY 836 LGAKDLPSBEDGRPR-NPVYKIYFLRDRSKNNKRRITVYKTLERKNNQITLYSPVHRE 894

Db 1078 IARGKLSKSDOSASADNPVKVLLPGRKVKSHKRRTREYVSSCAPENNVLEYO-VAPHT 1136

QY 895 FREMELEITLMDQARVREESEFLEGLIYL-ETALLDDEPHWYKLTQTHVSSLPLRPPS 953

Db 1137 LNTMFLFVLCYDQ--RDVDLPLGNVQITLADKSAINTGRMYPLQ-----GSTDQPIRQ 1190

QY 954 PVLPRRQLHGESPTRLQSRKISDSEVSDYDCEDGVGVVSDYVHNHGRDLOOSTVLEBQ 1013

Db 1191 HYM-----NCTSL-----QIPSI 1204

QY 1014 WMSHNCSPSSGSPHRNDVIGRTSMSPSPAPRQANVQGHGRTP-----ATGHYN 1063

Db 1205 AAASKHSAVAHNHH-----NHNYSFVPSITLYPKGVGTGRHDPKRVNHAFTENYN 1253

RESULT 6

T34318

hypothetical protein K03A1.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34318

R:Nhan, M.

submitted to the EMBL Data Library, November 1995

A:Description: The sequence of C. elegans cosmid K03A1.

A:Reference number: Z21505

A:Accession: T34318

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-547 <NHA>

A:Cross-references: EMBL:U41625; PIDN:AAA83324.1; CESP:K03A1.3

C:Genetics:

C:Gene: CESP:K03A1.3

C:Introns: 33/3; 50/2; 95/1; 118/3; 181/2; 293/3; 342/1; 371/3; 390/2; 451/1; 484/2

| | | | |
|----|-----|--|-----|
| Db | 143 | BE0P0K7ISSGSPSP0PLTD0-NVPEP00RAEP0DKMTPTPY0NN0P0R0G0M0P0NH | 201 |
| Qy | 264 | QIASMPDSDRRKSPSVSRD0NRRRYE0SEBEREDU5Q0YVPSD0TMRSPSDYADRSD0REP0 | 323 |
| Db | 202 | NOT00ENFM0N0NS0N0P0R0N0H0N0M0M0P0N0H0N0Q0N0H0G0ANN0Q0NNRR0M0Q0PM | 261 |
| Qy | 324 | FYEEF0G1LNDYSN0RGRHRSKEY1VDEDEDVSESDXEY0R0REEEY0ARY-SDP1N1ARY | 382 |
| Db | 262 | SON0M0N0N0M0N0-----N0Q0SH0N0M0T0N0P0N0G0P0N0Q0RTD0SR1M0K0 | 310 |
| Qy | 383 | PYK0P0YEE0M0R1HA0EY5RAR0EHRHSDV5LANAELDSR1SLRM-DPDSR0RSYERR | 441 |
| Db | 311 | TP0Q0P0S0Y0NNVGA---A0H0H0H0G0ED0H0Q0M0M0Q0RTDNNR0M0E0NT0G0G0M0NR0 | 367 |
| Qy | 442 | AAME-----N0R-----SYSMERT0EA0G05SY0P0RT5NHSPTPRR | 478 |
| Db | 368 | PSLE0TTPM0NKY0NHVEDD0G0M0R0P7FYTG0NSE0N0R0P0D0G0M0G0SQ0NN0N0Q0N0 | 422 |
| Qy | 479 | SP1PDRD0MRADSLK0H0HLDPSSA0VAKTK0REK0METL0R0NDSLSSDQ0SEYVR-P0PR | 537 |
| Db | 423 | -----NNBNLKR0N0FVS0RTEBDY0ASSNFE0E0K0R-----NNS0Q0S0M0T0G0V0R0C0STD | 472 |
| Qy | 538 | PHKSKKGGKM-R0Y-SLSSSEEL-----ASTP0EYTSDD0VELLES-ES | 577 |
| Db | 473 | DHLN0VKNR1R0L0R0SMSS0EED1I0AGG0NTL0K0MST5AV0ASG0K0TAF0DDMGAS0V0R | 532 |
| Qy | 578 | VSEKDS0KG 587 | |
| Db | 533 | ISE0CNS0ERG 542 | |

```

RESULT      7
S44644
Hypothetical protein F37A4.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 24-Sep-1999
C:Accession: S44644; T16287
R:Fulton, L.
Submitted to the EMBL Data Library, February 1994
A:Description: The sequence of C. elegans cosmid F37A4.
A:Reference number: S44638
A:Accession: S44644
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1021 <FUL>
A:Cross-references: EMBL:U00032; NID:9458958; PID:9458965; PIDN:AAA50635.1; CESP:F37A4
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP.F37A4.7
A:Introns: 74/1; 101/2; 153/2; 198/3; 234/3; 344/2; 462/1; 519/1; 592/3; 616/1; 741/1
C:Superfamily: protein kinase C C2 region homology
F:736-851/Domain: protein kinase C C2 region homology <KC2A>
F:876-991/Domain: protein kinase C C2 region homology <KC2B>

```

```

Query Match      4.38; Score 358; DB 2; Length 547;
Best Local Similarity 23.08; Prid. No. 1,4e-10;
Matches 140; Conservative 99; Mismatches 225; Indels 146; Gaps

QY 33 TEERKILAVMDQKKEEKEKESVLKIKKEHKAPQWMEPFSGITELVNVNLOPOOKP 92
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 24 TQNNGYAIKRFVKRQKBEAKE-----"Q-----ISKASEELSDLKQ 63

QY 93 NEKEPQRTLHQPFMYKEQYVKMKGEESQOOQEQKGDAPTCGICHTKTFAPAGCGHNSYQ 152
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 TERKETSK-----KLVTQ-----DQAIOGCKRTFPAIGIGHKCYCQ 102

QY 153 TKFCARCGGRVLSKNVVMVYCNICRQOELITSGAMFVNSGNSNTLOQPDQKVPGLRN 212
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 103 LRSARCGGGRQSK-NNAIMVCSLQCRQOOLATLQKWF-----QP----- 142

QY 213 EEAPOEKAKLHEQPOFQAGADISVAPEVKGRAHGLTROTDTIK-----NSGVK--H 263
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Query Match      4.18; Score 338; DB 2; Length 1021;
Best Local Similarity 20.9%; Pred. No. 3,1e-09;
Matches 197; Conservative 139; Mismatches 342; Indels 266; Gaps 43;

QY 90 KQNEKEPQRLHQOFEMFKPQVKM--GEESQOOOEQK-----GDAP 130
      : : : : : : : : : : : : : : : : : : : : : : : :
DB 114 RSPNTSAQCGSITAAE--QEHICKVLAKAESKSEQCRIGKMYDLKEMRRATNGV 171
      : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 T-QGICKRTP---ADQCGHNGSYCQKFCQR-GGGR---VSLRBNKV--MWVCNCRK 179
      : : : : : : : : : : : : : : : : : : : : : : : :
DB 172 THQLLCH-TEFGLLASASYAMCYDCRKYVQCRNGCVETTDVNTGKQYETVLCKTCS 230
      : : : : : : : : : : : : : : : : : : : : : : : :
QY 180 QQEL-TKSGAMFYSSNTLQOPDQKVPRLG--RNEAPQEKKALLHQPOFGAGP- 234
      : : : : : : : : : : : : : : : : : : : : : : : :
DB 231 AREVLAKMSGAWFYK-MPEQRDPDDRLPYVPVTTNGTLPNASSA-----ATPLSGPG 285
      : : : : : : : : : : : : : : : : : : : : : : : :
QY 235 ----DISVPAVE-----KRAHGLTRQDITKNSGVKHQIASDMPDRKRS----- 276
      : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 286 AGPDMTPMSTSSCOMTTPKMASPGVCSNPGLOMNGP-----TSPLENGTRRRTGEGIE 341
OY 277 -PSVRDQRRARRRQSEEREDYSQYVP-----SDGTPRSPEDYADRRSQRPQFEYEEGHL 331
Db 342 PRSSSDGSEVQSCVPKRALNKTTPVGSATTSAPRPPTSTTPSRRE-----A 392
OY 333 NYRDSNRGRHRSKEYIVDEDEVESRDEYERORREFEYQARYSDPNLARYPYAKPOPEE 391
Db 393 NMERRSRHTHAHANRLYSTDDDDSSPESRPTSRSTS-----PRHSL----- 434
OY 392 QMRTHAEVSRARHERHSDVSLANAELEDSRISLILMRDRSRQSRVSERAAAMNORSYS 451
Db 435 -----ATPSSYADTCHD-PTSLPDADTRSIDSGVQSDHSPQDS-----GLTCS 478
OY 452 MERTRAQOQSSVPORTSNIS--PPTPRRSPILPLDRPDMRADSLRKOHLLDPSSAVRKTG 510
Db 479 SSSLPLPQQQASH-----DHHSGGTPRR-----ISNP-----RTS 511
OY 511 REKMETMLRNDLSLDQSESVPRPPRPKSKKGKMKROYSLSSSEELASTPEYTSDD 570
Db 512 RV-----AQSASGTSIVTPPP-----ISSRTSPDNCSPLNV----- 545
OY 571 VELSESYSKGSQSGKRT--SEQVGLSDSNTRSEROKKMYGGHLEEDLEKSEPOI 629
Db 546 --MEKSSSSASTASSGGNRVGSAPVNLNHHAMHNNQ-----NHNDIKKLISQTSR 596
OY 630 KDSGVDTCSSTTNEHSHSDKHPVTWQPSKDGRLIGRILNKLKDGSVPRDSGAMIG 689
Db 597 AESPLAASSSFLSPDDDTQKN-----RRDYGVRVNSLQRLTSLDVAAP----- 643
OY 690 LKVVGGKMTESGRCAFTVYKKGSLADYGHILRPGDEVLEWNGRLLOGAFEEVYNTIL 749
Db 644 -----VAPISKMN-----GH-----IV 655
OY 750 ESKPEPQV---ELVSPRIDIDIPRIDSTH--AQLESSESSSFESQKMDRPSISVTPMSP 804
Db 656 SSEPSTSTNOMHTSYPIPTVYVPEEKEKALTASTESSEPCGAVIEPLDIDENEP 715
OY 805 -----GMLRDV-----PQFLSGQISIKLMPDKVGHQILVITLAKDLPREDG 847
Db 716 KHASRRGDCVGRVFCILCFSRKORSISITLTLTYHSAQRKILMHLIRAKNLKAMDSN 775
OY 848 RPRNPYKIFLDRSRKKNRRTKTKVKKLTLEPKNNQPIFYSPVHNRERERLETLMDO 907
Db 776 GFSDPYKFFLLPENTATYKLTSTKTEKTLNPEWMEKMYSGITEDDEKILRTVLDK 835
OY 908 ARVREESEFLGELILEETALDLDEPHMYKLGTHDVSSLPRLR 951
Db 836 DRI---GSDFLGERTALK-KLNDENMKKFNLYLE--SALPYPQ 873

RESULT 8
A45973
trichohyalin - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
C:Accession: A45973
R:Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M.
J. Biol. Chem. 268, 12164-12176, 1993
A:Title: The structure of human trichohyalin. Potential multiple roles as a functional E
ed (cross-linking) protein.
A:Reference number: A45973; MUID:93280194
A:Accession: A45973
A:Molecule type: DNA
A:Residues: 1-1898 <LEE>
A:Cross-references: GB:LO9190; NID:9292835; PIDN:AAA5582.1; PID:9292836
A:Note: authors translated the codon AGG for residue 1714 as Pro
C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh
Covalent modifications to this protein include conversion of arginine to citrulline and
C:Genetics:
A:Gene: GDB:THH
A:Cross-references: GDB:136223; OMIM:190370
A:Map position: 1q21-1q21

```

C:Superfamily: trichohyalin; calmodulin repeat homology
 C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
 F:49-81/Domain: calmodulin repeat homology <EP2>

```

Query Match          3.7%; Score 307; DB 1; Length 1898;
Best local similarity 17.4%; Pred. No. 2,3e-07;
Matches 245; Conservative 213; Mismatches 514; Indels 434; Gaps 53;

OY 19 QPPQPEMPDLSHLTEEE--RKTLAVMDRQKKEEEKDOSVLTKEENKAQPIQWTFPSG 76
Db 264 EPQQRLEQEE-----EEQRLKLEERQELRERQEEQOQRLRREQQLRRQEE----- 312
OY 77 ITELNVNVLQPOQKQRPNEKPEQTKLHQPEMYKEQYKKGSESOQOQ---EQKDAPTCG 133
Db 313 -----ERREQOERREQOERREQOERREQOERREQOERREQOERREQOERREQEE----- 358
OY 134 ICHKTKFADCGCHNSYCOTKFCARCGHVSLSRKNVWVCNLCKRQOETLTKSGAMFYN 193
Db 359 -----REEQQL----- 364
OY 194 SGNNTLQDPQKVPQGLNBEARQEKKAHLHEQPOFQAGBGLSVAVEGRAHGLTRQD 253
Db 365 ---RREQEEERREQQLRREQEEERREQQLRREQQLRRQD-----QLRREQQLRREQ 412
OY 254 TIKNSGVKHOIASDMPDRKRSPSVSDONRRYEQSEER--EDYSQYVPDGTMPSPSD 312
Db 413 QLRREQQLRREQQLRREQQLRREQQLRREQQLRREDEENHDKHDERREQQLKRRQEE 472
OY 313 YAD-----RRSQREPOFEYEEPH--LNYRDSNRGRHRSKEYIVDEDEVE 355
Db 473 RRDWLKREETERHNEQRRKQQLKRDQEEERRERWLKEEERREQOERREQQLRRQEE 532
OY 356 SRDE-YEQRREREYQARYSDPNLARYPYKQRYEROMTHAEVSRARHERHSDVSLA 414
Db 533 RREQRLKROEEERLQORLSEQQLR--EQEERLQQLKREEEKLDEQRQRL--- 586
OY 415 NAELEDSRISLMDRPSRVSERRAAMENORSYMERTRERAGQSSVTPORTSNHSP 474
Db 587 KRQOERROQLKREERERROQLKRDQEEERLQQR-LKREVEYRLQOERRDRLKREPE 645
OY 475 TPRRSPPLDRPDMRADSLRKQHLLDPSSAVRKTKEKMETMLRNDLSLDQSESVRPP 534
Db 646 EERRH-----ELKSEEEERRH--DQLRQOERREQQLRRE-----EEERLEQR 690
OY 535 PRPRHKKKGKMKROYSLSSSEELA-----SPEYTSDDVLESESESEKQDSQK 588
Db 691 LKREHEER---RQELAEEDQOARERIKSRIPKM---QWQLESEA-----DARQSK 737
OY 589 -RKTSEQVGLSDSNTRSEROKKMYUGHSLDEDEWSEFQIKDSGVDTCSSTLNEBHS 647
Db 738 VILEAPQAGRAEAPQOBEKRRR-----ESELQWQ-----EERRA 772
OY 648 H-----SDKHPVTWQPSKDGRLIGRILNKLKDGSVPRDSGAMIGKVVGGKMTESG 701
Db 773 HROQOEEERQDRFTWQAEKSEKRGQRLSAR-----PRLEQERQRLKEEQRQERQR 827
OY 702 RLCAFLTKVKKSLADYGHILRPGDEVLEWNGRLLOGATEFEVYNTILSEKPEQVELV 761
Db 828 FL-----PEEKEEQQR-----QREBEKELOF 851
OY 762 SRPIGDIRIPDSTHAQLESSESSSFESQKMDRPSISVTPSPMSPMLRDVYPLSGQLSK 821
Db 852 -----LEE-----FQQLRERRAQOQLEEEEDGLOEEOERRRQOERQDQ 889
OY 822 LWDKVGHOILVITLAKDLPREDGPRNRYVYILFLPDSDNKRRTKVTKTLEPKW 881
Db 890 KW-----RWOL-----EEKRRRRHT-LYAKPALQOQLKQOQLQOEEEBELQ 931
OY 882 NQPIFYSPVHNRERERMLLETLMDOARVRESESEFLGELILEETALDLDEPHMY---- 937
Db 932 REF---REKRRRQOERQOYR-----EEBQLOEEDQLLRE---ERKRRRQOERERQYRDX 981

```

[illegible]

| | | | |
|----|------|---|------|
| OY | 309 | SFDVADRSSRQPPCYEEEGCHLNTYD-----SNRGHSHSEYIYDDDEVEDRD---- | 358 |
| Dd | 233 | -BSEH RMWGGHGSESRHESHSHNSDRGKASSSSSGGHHNRQSSGDQERREVEETEE | 291 |
| OY | 359 | --EYEROREEEEOAUYRSDPDLNARYPVKPQPYEQOMRIAEYSRAHERRHSDVLANA | 416 |
| Dd | 292 | IYEEERHYAYEEERGARRS-----HYEESH--RHMSRGPGS----- | 326 |
| OY | 417 | ELEDSTISLLRMD---PRSQRSVSEERRAMENORSTSMRTLEAOG---QSTPOTSN | 470 |
| Dd | 327 | --QSSRASDSKODGHTRGSKERTRETETEADERNHGIEERREGORLGFVEESS--QRQES | 382 |
| OY | 471 | HSPPTRRSPPIRLDPDMRMDALRKONHLDPSSAVYKTREK-----M | 514 |
| Dd | 383 | HSD-----KGDSWRNDYEHSSTEPKRGAELEARDPAEEOGKHQASRGL | 427 |
| OY | 515 | ETWLANDSL----SSDOSESVRP RP RHKKSKG-GKMROVSLSSESSEELASTPEXTSC | 568 |
| Dd | 428 | EASRORESLIRGRKTTSSOSDOLR-----GLGESGATEVERARHNFTGEYEYDT | 476 |
| OY | 569 | DVELESESVSEKDGOKGKRKTSQ---GVLSNSTBREOKKRYAYGGHGLEEDLE- | 623 |
| Dd | 477 | PHAE-BSHOREESPTEHRGDEHNEHDYOANCMBEEBGHYOAOOGEOAMONEEEEEEE | 535 |
| OY | 624 | -----WSEPOIKDSDGVDTCSTTLNAB-----HSHSDKNHPYTWOPSKD--GDR | 664 |
| Dd | 536 | EDLFLEAMRSSPAYK-----AKSTLKENDPASFLPTRHQPQPSRMHNSNKQAWRDQ | 588 |
| OY | 665 | LIGRIILNKLKQGSVPBROSGAML---GLKVYGCKMTESRCLEFITVKVGKSIALDTVGH | 721 |
| Dd | 589 | LNTAADKSOELEFFSLHSHSRAKFRPALPYVNKKANLSSFESSPARIDELISLOGBGN | 648 |
| OY | 722 | --LRPDEVL-----EMNGRLGATGFEELYNIILESKREPOVELVYVRPGIDIPRIST | 775 |
| Dd | 649 | ASIRBPPEVLDRGOANGKYSSGSGALSERSOHLPTOQEABRKIGTT-PLG--RVSSQ | 704 |
| OY | 776 | HAOLESSSSSEF-----SOKMDPSISV---TSPWSGMLDVP | 811 |
| Dd | 705 | RUSVPYSOKPFQCSPLGTAPVLGSGOSSRPLOPORVLRHSSIIIGSSSPSGSIYTPVR | 764 |
| OY | 812 | QPLSGQLSTIKLWDXYGHOLIYLILCAKDLPNSE-DGRPNRPYUKITFLDR----- | 862 |
| Dd | 765 | Q-----KONTYPARE-----PKA-RPLEERDGGTGTVTALRPLDNHSSGGCTT | 810 |
| OY | 863 | ---SDKNKRRTYTKVKTLEPKNN---QTFIYSPVRYREFREMLITLMDQARVEEE | 914 |
| Dd | 811 | PNSITDKPOLLGLESISKLESDPRKXVEKEPRENE-----EEOGREED | 858 |
| OY | 915 | SEFLGETLIELEALLDERHNMYKLOTHDVSLCPLRPSPYLL-RQNLGESPTRRLQ- | 971 |
| Dd | 859 | MODYDEI-VELEEEOSSAAE---SFERRPLRSLRDKNSSPIKSALARPTGKHTGRAVOE | 914 |
| OY | 972 | -RSKRIDSQVSUQDCEDGVGYVDYHNHNRDLOQSLSLYPEOQMSSNHCSPPSGSPIRYD | 1030 |
| Dd | 915 | ANSVROSPTPRNAADBOQRVV-----EGR-----STRVIITSSPSTHTESP--- | 958 |
| OY | 1031 | VIGTRSWSPSAPPORNVBQAGRTGATGHNYTISHMDRHNVADHYSSDRDRCEAAD | 1090 |
| Dd | 959 | -----VHSPAGSEBIRE-----NATGMEDYQDEBEONHELDTVDMDIMDYMD-D | 1003 |
| OY | 1091 | KOPYHNS-----RSTEOBRLLETITTRRSRSSRPTNLMRSMPSLMTGKSAPS--PA | 1141 |
| Dd | 1004 | LEMYESHPAPARAQAARPAPRRPRLSPRGKHQROOOOQOELOOQVAAAARPORAOPR | 1063 |
| OY | 1142 | LSRSHRTGSVQSP-----SSTPGTGRROROLRPLPKG-----TLERSAMDIEER | 1188 |
| Dd | 1064 | LKRP---TSTLEMSPYLRSGVSLFGCGRKKSLPRPLRPNSKNINITYPLQPLYSPLPVSSF | 1120 |
| OY | 1189 | NROMKLANKYQAVAGSPRLBODYHSKTRSGMDPHRG-ADTVSTKSSSDSYSDVASVETS | 1247 |
| Dd | 1121 | GQQ-----QOOOQOOOSSPSTJSTOSH-----HNGCASGTGASIGASSSTVSASAATTTTT | 1168 |

A:Cross-references: EMBL:U08913; NID:93850884; PID:93850885; PIDN:AACT2432.
A:Experimental source: Strain Balb/C
C:Genetics:
A:Gene: P2P-R
C:Function:
A:Description: involved in hnRNP association and Rb1 binding
C:Superfamily: RING finger homology
A:57-107/Domain: RING finger homology <RRN>

[illegible]

```

OY      887  ---YSPVHREPR-----EMLLETTIMDQAVREES--EFLGELLELEALL 930
           |||
           |||
Db      1249 DSRDDSSKKRDERGELARRKDSPPRGKESLSGOKSLREERDLPPKGAESKKSNSSPPR 1308
           |||
OY      931 DDEPHWYKQTHDVSSLPYRPSRYLPYRQDHGESPRLR---ORSKRIDSEVYDCE 987
           |||
           |||
           |||
Db      1309 DKPRDHK-----APYETKRCEETKPYDKLSGKEREKNAEA----- 1346
           |||
OY      988 DGVGVVSDYRHNGRDLQSSSTLSVPEQVMSSNHC--SPSGSPHRYDV-IGTRSMSPSAPP 1044
           |||
           |||
           |||
Db      1347 -----RNGKD--SSGGLRP-----CLNRPDLMEKELAVGVQVEK-SAVKPK 1385
           |||
OY      1045 PORNVQGRHGRTRANGHYNTISMMDHNRWMDHYSSDRDPCSAADROPYHNRKSTE--- 1101
           |||
           |||
           |||
Db      1386 PQ-----LSHSR-----LSSDLRETLEAEPRDYNESDSNNY 1420
           |||
OY      1102 -----QRPLERTTTRSS-----ERP--DTLWMSMPSLMTMGSRSPSPAL 1142
           |||
           |||
           |||
Db      1421 SVKEEBAVASISDLKETEKTTEKKESSLVYATASOPGADRQSOSSPSVSRSHSPSGSQ 1480
           |||
OY      1143 SRSHPTGSVQTSPPSTPGTGRGROLPOLPRKGLTERSAMDEERNRQKMLINKYQVAG 1202
           |||
           |||
           |||
Db      1481 TSHSSSSASAGQDBK-----KKKKKKKKKKKKKKKKKKHAG 1522
           |||
OY      1203 SUPPLEQDYHNSKTRSGMDPRHGDADTVSTSSDSVDVSVAVS 1244
           |||
           |||
           |||
Db      1523 AGDVYKSQKHKK-----KKKKKKKKDEKEKDDQKVRST 1559
           |||

```

RESULT 12
T20531
hypothetical protein F07A11.6a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T20531; T27776
R:Palmer, S.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z19287
A:Accession: T20531
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2526 <WIL>
A:Cross-references: EMBL:Z66511; PIDN:CAB54210.1; GSPDB:GN00020; CESP:F07A11.6a
A:Experimental source: clone F07A11
R:Gajadsky, S.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z20417
A:Accession: T27776
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2526 <WIL>
A:Cross-references: EMBL:Z65904; PIDN:CAB54501.1; GSPDB:GN00020; CESP:F07A11.6a
A:Experimental source: clone ZK20
C:Genetics:
A:Gene: CESP:F07A11.6a
A:Map position: 2
A:Introns: 36/2; 92/2; 182/2; 272/3; 344/2; 426/2; 490/1; 541/2; 700/3; 770/3; 1286/3

```

Query Match          3.2%; Score 266; DB 2; Length 2526;
Best Local Similarity 17.9%; Pred. No. 3.5e-05;
Matches 298; Conservative 235; Mismatches 592; Indels 540; Gaps 66;

QY 180 QOEILTKSAMVYNNSSNTLQOPDPQKVPGLRNEAPKAKKLLHQPQFGAPGDLSP 239
      ::::|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 908 RHDITTSKSP---SNAINALQARSQMTSGDEKSKAP-----STPVRDAGSDLVAQ 956
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 240 AVEKGRAHILTEPDQITKRNKSGVKKHIASDMP---SDPKRSPY-----SPQDN 284
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 957 IMSNQNLGRLKRLPRLEKSSALQNTQNHQPHSNANSTFPTSTPSTQIQAAMKKDEKEKK 1016
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 285 RVEQSEEREDYQGVVPSDQITPSPSDVADRSSQNEPDVEEPGLNRYDSNRGRHRS 344

```

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2001, 16:00:09 ; Search time 29.05 seconds
(without alignments)
3318.146 Million cell updates/sec

Title: US-09-617-099b-1
Perfect score: 8285
Sequence: 1 MSALGPGRPAAPTAAASQP.....TRPASQSLSTGTPSYRS 1590

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_0601.*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT.*
23: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|-----------|--------------------|
| 1 | 1586.5 | 19.1 | 374 | AAW29640 | Human secreted pro |
| 2 | 1015.5 | 12.3 | 237 | AAAB34848 | Human secreted pro |
| 3 | 1010.5 | 12.2 | 237 | AAAB34847 | Gene 46 human secr |
| 4 | 624 | 7.5 | 134 | AAAB34849 | Human secreted pro |
| 5 | 566 | 6.8 | 128 | AAAB34850 | Human secreted pro |
| 6 | 420 | 5.1 | 126 | AAAB40725 | Human ORFX ORF489 |
| 7 | 417 | 5.0 | 86 | AAAB55848 | PDZ encoded domain |
| 8 | 417 | 5.0 | 86 | AAAB57639 | PDZ domain. Homo |
| 9 | 417 | 5.0 | 86 | AAAB58049 | PDZ domain. Homo |
| 10 | 307 | 3.7 | 1898 | AAAY30795 | A human trichohyal |
| 11 | 270.5 | 3.3 | 3266 | AAAB42491 | Human ORFX ORF255 |

| | | | | | | |
|----|-------|-----|------|----|-----------|--------------------|
| 12 | 258.5 | 3.1 | 2742 | 21 | AAAB23012 | Human APC protein |
| 13 | 258.5 | 3.1 | 2842 | 15 | AAAR63508 | Adenomatous polypo |
| 14 | 258.5 | 3.1 | 2843 | 16 | AAAB19922 | Adenomatous polypo |
| 15 | 258.5 | 3.1 | 2843 | 19 | AAAB76140 | Human APC protein |
| 16 | 258.5 | 3.1 | 2843 | 19 | AAAB76144 | Human APC protein |
| 17 | 258.5 | 3.1 | 2843 | 21 | AAAB23011 | Human APC protein |
| 18 | 258.5 | 3.1 | 2973 | 19 | AAAB76821 | Human APC protein |
| 19 | 258.5 | 3.1 | 2973 | 21 | AAAB70304 | Protein used in ca |
| 20 | 258.5 | 3.1 | 2973 | 22 | AAAB72782 | Transcriptional ac |
| 21 | 256.5 | 3.1 | 704 | 15 | AAAB57421 | Rabphilin-3A. Hom |
| 22 | 253.5 | 3.1 | 1341 | 15 | AAAB5657 | Human Actin L. pro |
| 23 | 253.5 | 3.1 | 2843 | 15 | AAAB5657 | Adenomatous polypo |
| 24 | 250.5 | 3.0 | 2843 | 13 | AAAB26052 | APC gene product 1 |
| 25 | 250.5 | 3.0 | 2843 | 18 | AAAB5392 | Human adenomatous |
| 26 | 250.5 | 3.0 | 2843 | 19 | AAAB38370 | Human adenomatous |
| 27 | 250.5 | 3.0 | 2860 | 15 | AAAB3507 | Adenomatous polypo |
| 28 | 248.5 | 3.0 | 2343 | 21 | AAAB12453 | Human HNRCP protei |
| 29 | 243 | 2.9 | 1596 | 18 | AAAB1347 | Rat tumour suppres |
| 30 | 235 | 2.8 | 2432 | 21 | AAAB5565 | Human homologue of |
| 31 | 230.5 | 2.8 | 2453 | 21 | AAAB12454 | HNRCP protein sequ |
| 32 | 230 | 2.8 | 2819 | 22 | AAAB35408 | Human OTCG27 gene |
| 33 | 226 | 2.7 | 2101 | 21 | AAAB49936 | Human Numa protein |
| 34 | 225.5 | 2.7 | 1780 | 19 | AAAB3863 | Human gravin polyp |
| 35 | 225.5 | 2.7 | 1780 | 21 | AAAB15380 | Human gravin prote |
| 36 | 222 | 2.7 | 2101 | 15 | AAAB47173 | Sequence of the in |
| 37 | 220 | 2.7 | 2115 | 21 | AAAB49937 | Human Numa protein |
| 38 | 217 | 2.6 | 917 | 21 | AAAB33920 | A Bcl-2 associated |
| 39 | 217 | 2.6 | 2192 | 18 | AAAB21732 | LexA/Numa fusion p |
| 40 | 217 | 2.6 | 2272 | 18 | AAAB21731 | GAL4/HA/Numa fusio |
| 41 | 212.5 | 2.6 | 1054 | 22 | AAAB68897 | Human HX2004-6 pro |
| 42 | 210 | 2.5 | 2213 | 20 | AAAY6079 | Human actin-bindin |
| 43 | 209 | 2.5 | 1239 | 20 | AAAB55931 | Human zc1 protein. |
| 44 | 208 | 2.5 | 2442 | 21 | AAAB77575 | Human cytoskeletal |
| 45 | 207.5 | 2.5 | 1064 | 22 | AAAB68898 | Human HX2004-6 pro |

ALIGNMENTS

| | | | | | |
|----------|-------------------------------------|-----------------------------------|---|--|--|
| RESULT | 1 | | | | |
| AAW29640 | AAW29640 standard; Protein: 374 AA. | | | | |
| XX | AC | AAW29640: | | | |
| XX | DT | 09-NOV-1998 (first entry) | | | |
| XX | DE | Human secreted protein C0618_1. | | | |
| XX | KW | Secreted protein: C0618_1; human. | | | |
| XX | OS | Homo sapiens. | | | |
| XX | FT | Key | Location/Qualifiers | | |
| FT | FT | Misc-difference 100 | /note= "encoded by WAA" | | |
| FT | FT | Misc-difference 269 | /note= "encoded by TGS" | | |
| FT | FT | Misc-difference 289 | /note= "encoded by YTA" | | |
| FT | FT | Misc-difference 372 | /note= "encoded by TYT" | | |
| FT | FT | Domain | /note= "predicted transmembrane domain is centered around amino acid 350" | | |
| PN | PN | WO9831802-A1. | | | |
| XX | XX | 23-JUL-1998. | | | |
| XX | XX | 21-JAN-1998: | 98WO-US01007. | | |
| XX | XX | 20-JAN-1998: | 98US-0010047. | | |

| | | |
|----|--|---------------|
| PR | 21-JAN-1997; | 97US-0072103. |
| PR | 16-JUN-1997; | 97US-0877035. |
| XX | | |
| PA | (GEMV) GENETICS INST INC. | |
| XX | | |
| PI | Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D; | |
| PI | Racle LA, Spaulding V, Treacy M; | |
| XX | | |
| DR | WPI: 1998-414101/35. | |
| DR | N-PSDB: AAV40485. | |
| XX | | |
| PT | New isolated polynucleotides and secreted proteins - obtained from | |
| PT | human foetal kidney, adult testes, placenta, adult brain and foetal | |
| PT | brain cDNA libraries | |
| XX | | |
| PS | Claim 22; Page 71-72; 104pp; English. | |
| XX | | |
| CC | This is the amino acid sequence of novel human secreted protein | |
| CC | C0618_1, as predicted from human adult brain cDNA clone C0618_1 | |
| CC | (see AAV40485). C0618_1 cDNA was isolated using methods which are | |
| CC | selective for cDNAs encoding secreted proteins, and was identified | |
| CC | as encoding a secreted or transmembrane protein on the basis of | |
| CC | computer analysis of the amino acid sequence of the encoded | |
| CC | protein. C0618_1 shows some sequence similarity to rat Rim, a | |
| CC | putative Rab3 effector. Novel secreted proteins (see AAV29636-45) | |
| CC | have been identified in human foetal kidney, adult testis, placenta, | |
| CC | adult brain and foetal brain, and can be obtained by expression in | |
| CC | recombinant host cells. The polypeptides may have e.g. nutritional | |
| CC | activity, cytokine and cell proliferation or differentiation | |
| CC | activity, immunostimulant or immunosuppressive activity, | |
| CC | haematopoiesis regulating activity, tissue growth activity, | |
| CC | activin/inhibin activity, chemotactic or chemokinetic activity, | |
| CC | haemostatic or thrombolytic activity, receptor/ligand activity, | |
| CC | antiinflammatory activity, cadherin or tumour invasion suppressor | |
| CC | activity, tumour inhibition activity, or other activities. | |
| XX | | |
| XX | Sequence 374 AA; | |

| | | | | |
|---------------------------|---|-------------------|------------|-------------|
| Query Match | 19.1% | Score 1586.5; | DB 19; | Length 374; |
| Match Local Similarity | 85.8% | Pred. No. 2e-103; | | |
| Matches 321; Conservative | 6; | Mismatches 6; | Indels 41; | Gaps 1; |
| QY 1258 | MSVQSERPGRNRKISVETSKMQRNROMGVSGKNILTKSTISIGDMCSLEKNDGSQSDTAVGA | 1317 | | |
| Db 1 | msvqserpggnkktisvtfckmgsrqmgisgkrmktstisgdmcslekdngdsqstavyt | 60 | | |
| QY 1318 | LGISGKKRRRSITGAKMAVIVGLSRKSRASASOLQTEGGGKKLRSTVQBSTETGLAVEMRN | 1377 | | |
| Db 61 | lgtsgkkrtrssalgaakmvaivglstrksrassqslgeaagxkrlrstvtqrstetglavemrn | 120 | | |
| QY 1378 | WMPTQASSESTNDGSNNVSSSGCNLIFPEPVRLASDQSEDFLDGCPADLVNGQTLATPAM | 1437 | | |
| Db 121 | wmtqtastesldgsumnsysegcnliffpvrlasdsqisdfldglpdaqivgrqlatpam | 180 | | |
| QY 1438 | GDIOVGMDKKGOLEVEIIRARGLVKRGSKTLPAPYKVVLLDNGVC----- | 1485 | | |
| Db 181 | gdiovgmddkkgqlveleirarglvvkpysktlpapykvvllndvgtlakkqgleveii | 240 | | |
| QY 1486 | -----IAKKKTIVARKTLEPLLYOOLLFEESPOGRV | 1516 | | |
| Db 241 | rarglvvkpysktlpapykvvllndvgtlakkktkvarktleplyqllsfeespgkv | 300 | | |
| QY 1517 | LQITVWDGYGMDKRSFNGVAQOILLDELELSNMVYIGWKLEPPSSLVVPITAPLTRASQ | 1576 | | |
| Db 301 | lqitlvwdygmddkrsfngvaqilldelelsnmvylgwklefppsslvvpitapltrrsq | 360 | | |
| QY 1577 | SSLESTGSPYSRS 1590 | | | |
| Db 361 | ssleestgspysrs 374 | | | |

| | |
|----------|--|
| ID | AAB34848 standard; Protein; 237 AA. |
| AC | AAB34848; |
| DT | 26-JAN-2001 (first entry) |
| DE | Human secreted protein sequence encoded by gene 46 SEQ ID NO:136. |
| KW | Human; secreted protein; diagnosis; cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic; cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder; wound healing; neurological disease; infectious disease; chromosome identification; chromosome 6. |
| OS | Homo sapiens. |
| PN | WO200058356-A1. |
| PD | 05-OCT-2000. |
| PX | 22-MAR-2000; 2000MO-US07535. |
| PF | 26-MAR-1999; 99US-0126511. |
| PR | 17-DEC-1999; 99US-0172413. |
| PX | (HUMA-) HUMAN GENOME SCI INC. |
| PA | Rosen CA, Ruben SM, Komatsoulis G; |
| PI | WPI: 2000-594639/56. |
| PX | Fifty nucleic acid molecules encoding human secreted proteins, useful in the prevention, treatment and diagnosis of cancer, immune disorders, cardiovascular disorders and neurological diseases - |
| PT | disclosure; Page 419-420; 425pp; English. |
| PS | The polynucleotide sequences given in AAC59966 to AAC60015 encode the human secreted proteins given in AAB34773 to AAB34822. AAB34823 to AAB34822 represent human secreted polypeptide sequences and proteins homologous to them, which are given in the exemplification of the present invention. Human secreted proteins have activities based on the tissues and cells the genes are expressed in. Examples of activities include: cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor; vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic; and cardiant. The polynucleotides and polypeptides are useful for preventing, treating or ameliorating a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. The polynucleotides are useful for chromosome identification. They are also useful as probes for diagnosing a disorder related to the female reproductive system, particularly breast and/or ovary cancer. They are also useful in the gene therapy of breast and ovarian cancer. Nucleic acids, protein, antibodies, agonists and antagonists from the present invention are useful in the diagnosis, treatment and prevention of cancer, immune disorders, cardiovascular disorders, wound healing, neurological diseases and infectious disease. AAC59967 to AAC59965 and AAB34772 represents sequence used in the exemplification of the present invention. |
| Sequence | 237 AA; |

| | | | | |
|---------------------------|--|--------------------|-----------|-------------|
| Query Match | 12.3% | Score 1015.5; | DB 21; | Length 237; |
| Best Local Similarity | 82.8%; | Pred. No. 1.4e+63; | | |
| Matches 197; Conservative | 19; | Mismatches 21; | Indels 1; | Gaps 1; |
| OY | EGCGKKLRSTVQRSTFEICLAYEMRNMMTRQASRRESTDGSMNSYSSEGNLFFPGVARLSDS | 1412 | | |

Db 1 esghkklkltqstetgmaeemrk-mvrgpserestdgsinsyssegnlifpgvrlgads 59
 QY 1413 QESDFLDGIGPAOLVGRQTLATPAMGDIQVGMMDKKGQLEVEIIRARGLVKPGSKTLPA 1472
 Db 60 qfsdflgdlgpaqlvgrqtlatpamgdlqgmedkkgqlevetirarstltqkpskstpa 119
 QY 1473 PYKVYVLLDNGVCIAKKRTKVAARKTLEPLVQQLSFESPSGCRVLIQIYWGDIYGRMDHKS 1532
 Db 120 pykvvylldngaciacakkttriarctldplyqgslyfdespgkqvlyvwdygrmdhkc 179
 QY 1533 FMGVAQIILDELEISNMVIGWFKLEPPSSLVDPTSAPILTRRASQSSLESSTGPGYSRS 1590
 Db 180 fmgyaqllleeldlssmviywkllfpsslvdpdltrrasqsslesstgpcirs 237

RESULT 3
 AAB34847 ID AAB34847 standard; Protein: 237 AA.
 AAB34847; AC AAB34847;
 DT 26-JAN-2001 (first entry)
 DE Gene 46 human secreted protein homologous amino acid sequence #135.
 KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
 KW neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antinflammatory; antitumor; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiac; gene therapy;
 KW cancer; immune disorder; cardiovascular disorder; wound healing;
 KW neurological disease; infectious disease; chromosome identification.
 OS Rattus norvegicus.
 PN W0200058356-A1.
 PD 05-OCT-2000.
 PF 22-MAR-2000; 2000WO-US07535.
 PR 26-MAR-1999; 99US-0126511.
 PR 17-DEC-1999; 99US-0172413.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Ruben SM, Komatsoulis G;
 DR WPI; 2000-594639/56.
 XX FIFTY nucleic acid molecules encoding human secreted proteins, useful
 PT in the prevention, treatment and diagnosis of cancer, immune disorders,
 PT cardiovascular disorders and neurological diseases -
 PS Disclosure: Page 418-419; 425pp; English.

The polynucleotide sequences given in AAC59966 to AAC60015 encode the
 CC human secreted proteins given in AAB34773 to AAB34822. AAB34823 to
 CC AAB34852 represent human secreted polypeptide sequences and proteins
 CC homologous to them, which are given in the exemplification of the present
 CC invention. Human secreted proteins have activities based on the tissues
 CC and cells the genes are expressed in. Examples of activities include:
 CC cytostatic; immunosuppressive; neotropic; neuroprotective; antiviral;
 CC antiallergic; hepatotropic; antidiabetic; antinflammatory; antitumor;
 CC vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic; and
 CC cardiac. The polynucleotides and polypeptides are useful for
 CC preventing, treating or ameliorating a medical condition in e.g. humans,
 CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities. The polynucleotides are
 CC useful for chromosome identification. They are also useful as probes for
 CC diagnosing a disorder related to the female reproductive system,
 CC particularly breast and/or ovary cancer. They are also useful in the gene
 CC therapy of breast and ovarian cancer. Nucleic acids, protein, antibodies,

CC agonists and antagonists from the present invention are useful in the
 CC diagnosis, treatment and prevention of cancer, immune disorders,
 CC cardiovascular disorders, wound healing, neurological diseases and
 CC infectious disease. AAC59957 to AAC59965 and AAB34772 represents sequence
 CC used in the exemplification of the present invention.
 CC
 SO Sequence 237 AA;

Query Match 12.2%; Score 1010.5; DB 21; Length 237;
 Best Local Similarity 82.4%; Pred. No. 3.2e-63;
 Matches 196; Conservative 20; Mismatches 21; Indels 1; Gaps 1;

QY 1353 EGGGKLRSTVORSENGTFLAVERMWMTRQASRESTDGSMNSYSEGNLIFPGVRLASDS 1412
 Db 1 esghkklkltqstetgmaeemrk-mvrgpserestdgsinsyssegnlifpgvrlgads 59
 QY 1413 QESDFLDGIGPAOLVGRQTLATPAMGDIQVGMMDKKGQLEVEIIRARGLVKPGSKTLPA 1472
 Db 60 qfsdflgdlgpaqlvgrqtlatpamgdlqgmedkkgqlevetirarstltqkpskstpa 119
 QY 1473 PYKVYVLLDNGVCIAKKRTKVAARKTLEPLVQQLSFESPSGCRVLIQIYWGDIYGRMDHKS 1532
 Db 120 pykvvylldngaciacakkttriarctldplyqgslyfdespgkqvlyvwdygrmdhkc 179
 QY 1533 FMGVAQIILDELEISNMVIGWFKLEPPSSLVDPTSAPILTRRASQSSLESSTGPGYSRS 1590
 Db 180 fmgyaqllleeldlssmviywkllfpsslvdpdltrrasqsslesstgpcirs 237

RESULT 4
 AAB34849 ID AAB34849 standard; Protein: 134 AA.
 AAB34849; AC AAB34849;
 DT 26-JAN-2001 (first entry)
 DE Gene 46 human secreted protein homologous amino acid sequence #137.
 KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
 KW neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antinflammatory; antitumor; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiac; gene therapy;
 KW cancer; immune disorder; cardiovascular disorder; wound healing;
 KW neurological disease; infectious disease; chromosome identification.
 OS Rattus norvegicus.
 PN W0200058356-A1.
 PD 05-OCT-2000.
 PF 22-MAR-2000; 2000WO-US07535.
 PR 26-MAR-1999; 99US-0126511.
 PR 17-DEC-1999; 99US-0172413.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Ruben SM, Komatsoulis G;
 DR WPI; 2000-594639/56.
 XX FIFTY nucleic acid molecules encoding human secreted proteins, useful
 PT in the prevention, treatment and diagnosis of cancer, immune disorders,
 PT cardiovascular disorders and neurological diseases -
 PS Disclosure: Page 420-421; 425pp; English.

The polynucleotide sequences given in AAC59966 to AAC60015 encode the
 CC human secreted proteins given in AAB34773 to AAB34822. AAB34823 to
 CC AAB34852 represent human secreted polypeptide sequences and proteins

CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Examples of activities include:
CC cytostatic; immunosuppressive; neutrotropic; neuroprotective; antiviral;
CC antidiabetic; hepatotropic; antidiabetic; antiinflammatory; antilucer;
CC vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic; and
CC cardiant. The polynucleotides and polypeptides are useful for
CC preventing, treating or ameliorating a medical condition in e.g. humans,
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. The polynucleotides are
CC useful for chromosome identification. They are also useful as probes for
CC diagnosing a disorder related to the female reproductive system,
CC particularly breast and/or ovarian cancer. They are also useful in the gene
CC therapy of breast and ovarian cancer. Nucleic acids, protein, antibodies,
CC agonists and antagonists from the present invention are useful in the
CC diagnosis, treatment and prevention of cancer, immune disorders,
CC cardiovascular disorders, wound healing, neurological diseases and
CC infectious disease. AAC59957 to AAC59965 and AAB34772 represents sequence
CC used in the exemplification of the present invention.

XX Sequence 134 AA;

Query Match 7.5%: Score 624; DB 21; Length 134;
Best Local Similarity 85.1%: Pred. No. 2e-36;
Matches 114; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

OY 825 DKVHOLIVTLGAKDLPESREDPRNPYVKIYFLPDRSDKNKRRTKYKTKLEPKMNQOT 884

DB 1 dkvghqlivmqlgtdlprdygrpnpykmyflpdrskskrtktvkllepwnqt 60

OY 885 FIVSPVHRFRERMLELTLDQARVRESEFLGEILIELELTALIDDPHWYKLOTNDV 944

DB 61 fvyghvhrdfrermltltvwdqprvgeeseflgeilietaltliddphwykltqhde 120

OY 945 SSLPLRPSPYLPR 958

DB 121 sslplrpqpslpmr 134

RESULT 5
AAB34850
AAB34850 standard; Protein: 128 AA.

XX AAB34850;

DT 26-JAN-2001 (first entry)

DE Human secreted protein sequence encoded by gene 46 SEQ ID NO:138.

XX Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW neutrotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antilucer; vulnerary; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiant; gene therapy;
KW cancer; immune disorder; cardiovascular disorder; wound healing;
KW neurological disease; infectious disease; chromosome identification;
KW chromosome 6.

XX Homo sapiens.

PN WO200058356-A1.

PD 05-OCT-2000.

PF 22-MAR-2000; 2000WO-US07535.

PR 26-MAR-1999; 99US-0126511.

PR 17-DEC-1999; 99US-0172413.

PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Ruben SM, Komatsoulis G;

XX WPI, 2000-594639/56.
DR Fifty nucleic acid molecules encoding human secreted proteins, useful
XX in the prevention, treatment and diagnosis of cancer, immune disorders,
PT cardiovascular disorders and neurological diseases -
XX Disclosure; Page 421; 425pp; English.

XX The polynucleotide sequences given in AAC59966 to AAC60015 encode the
CC human secreted proteins given in AAB34773 to AAB34822. AAB34823 to
CC AAB34852 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Examples of activities include:
CC cytostatic; immunosuppressive; neutrotropic; neuroprotective; antiviral;
CC antiallergic; hepatotropic; antidiabetic; antiinflammatory; antilucer;
CC vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic; and
CC cardiant. The polynucleotides and polypeptides are useful for
CC preventing, treating or ameliorating a medical condition in e.g. humans,
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. The polynucleotides are
CC useful for chromosome identification. They are also useful as probes for
CC diagnosing a disorder related to the female reproductive system,
CC particularly breast and/or ovarian cancer. They are also useful in the gene
CC therapy of breast and ovarian cancer. Nucleic acids, protein, antibodies,
CC agonists and antagonists from the present invention are useful in the
CC diagnosis, treatment and prevention of cancer, immune disorders,
CC cardiovascular disorders, wound healing, neurological diseases and
CC infectious disease. AAC59957 to AAC59965 and AAB34772 represents sequence
CC used in the exemplification of the present invention.

XX Sequence 128 AA;

Query Match 6.8%: Score 566; DB 21; Length 128;
Best Local Similarity 82.0%: Pred. No. 2.3e-32;
Matches 105; Conservative 13; Mismatches 10; Indels 0; Gaps 0;

OY 1446 DKRGQLEVEIIRAGLVYKPGSKTLPAHYKVVYLLDNCVCIAKKTKYAKRTLEPLYOOL 1505

DB 1 dkkqglevevirarsltcqgskstpapykvyllengaciacakkkrtlartklldlyqqs 60

OY 1506 LSPFESPQGRVLOIIVMGDVCVRMDKSPMGVAOITLDELISNMVIGMFKLEPPSSIVDP 1565

DB 61 lvidespqgkvlygvwdygrmdhkcimgvaqllleeldlssmvygkkltpssivdp 120

OY 1566 TSAPLTPRR 1573

DB 121 tlpltprr 128

RESULT 6
AAB40725
AAB40725 standard; Protein: 126 AA.

XX AAB40725;

DT 08-FEB-2001 (first entry)

DE Human ORFX ORF489 polypeptide sequence SEQ ID NO:978.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antiparasitic; antiparkinsonian; neutrotropic; neuroprotective;
KW anticonvulsant; osteopathic; antiallergic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antihemmatic; antihydro;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;

RESULT 8
AAB57639 standard; Protein; 86 AA.
ID AAB57639 standard; Protein; 86 AA.
XX AAB57639;
AC
XX
XX
DF 12-MAR-2001 (first entry)
XX
XX
DE PDZ domain.
XX
XX Endothelial cell; haematopoietic cell; PDZ domain protein;
KW PL domain protein; leukocyte activation; synapse formation;
KW transmembrane neurotransmitter receptor; autoimmune disease;
KW transplantation rejection; inflammation; allergy;
KW inflammatory bowel disease; ulcerative colitis; ileitis; psoriasis;
KW asthma; atopic dermatitis; atherosclerosis; cancer; infectious disease;
KW ischaemia; vasculitis; Crohn's disease.
XX
XX Homo sapiens.
OS
XX
XX WO200069897-A2.
PN
XX
XX 23-NOV-2000.
PD
XX
XX 12-MAY-2000; 2000WO-US13166.
PF
XX
XX 14-MAY-1999; 99US-0134114.
PR 14-MAY-1999; 99US-0134117.
XX 14-MAY-1999; 99US-0134118.
PR 21-OCT-1999; 99US-0160860.
XX 29-OCT-1999; 99US-0162498.
PR 13-DEC-1999; 99US-0170453.
XX 14-JAN-2000; 2000US-0176195.
PR 14-FEB-2000; 2000US-0182296.
XX 11-APR-2000; 2000US-0196460.
PR 11-APR-2000; 2000US-0196527.
XX
XX (ARBO-) ARBOR VITA CORP.
PA
XX
XX Lu PS;
PI
XX
XX WPI; 2001-025003/03.
DR
XX
XX New inhibitors of binding of a PDZ protein and PL protein for
PT inhibiting T cell-mediated response by hematopoietic cells, or for
PT treating diseases characterized by inflammatory and humoral immune
PT responses, e.g. inflammation, cancer -
XX
XX
XX Disclosure; Page 36; 139pp; English.
PS
XX
XX The present invention relates to a method for modulating a biological
CC function of an endothelial cell or haematopoietic cell, comprises
CC introducing into a cell an antagonist that inhibits binding between a
CC PDZ domain protein and a PL domain protein to result in inhibition of
CC leukocyte activation. The present sequence is a PDZ domain. PDZ domains
CC of proteins are named after three prototypical proteins: PSD95, PDZ domain
CC Drosophila large disc protein and Zonula Occludin 1 protein. PDZ domain
CC proteins are involved in synapse formation by organising transmembrane
CC neurotransmitter receptors through intracellular interactions. The
CC inhibitors identified by the present invention can be used to treat a
CC disease mediated by haematopoietic cells, e.g. autoimmune disease,
CC inflammation, allergy (e.g. drug allergies), inflammatory bowel diseases,
CC ulcerative colitis, ileitis, psoriasis, respiratory allergic diseases
CC (e.g. asthma), atopic dermatitis, autoimmune diseases (e.g. rheumatoid
CC arthritis, multiple sclerosis, insulin-dependent diabetes, Hashimoto
CC thyroiditis, osteoarthritis), atherosclerosis, cancers, infectious
CC diseases (e.g. viral infection), ischaemia, vasculitis and Crohn's
CC disease. The inhibitors can also be used to prevent transplantation
CC rejection of a solid organ transplant.
XX
XX Sequence 86 AA:
SQ

Query Match 5.0%; Score 417; DB 22; Length 86;
Best Local Similarity 100.0%; Pred. No. 3.7e-22;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 682 RDSGAMGLKVVGGKMTESGRICAFITKVKKGLADTVGHLRGGDEVLEWNGRLTGATP 741
DB 1 rdsгамlglkvvggkmtesgricafitkvkksladtvghlrpgdevlewngrltgatf 60
OY 742 EEVYNIIIESKPEPEVELVSR 763
DB 61 eevyniileskpepevelvsvr 82
RESULT 9
AAB58049 standard; Protein; 86 AA.
ID AAB58049 standard; Protein; 86 AA.
XX
XX AAB58049;
AC
XX
XX 12-MAR-2001 (first entry)
DF
XX
XX
DE PDZ domain.
XX
XX Endothelial cell; haematopoietic cell; PDZ domain protein;
KW PL domain protein; leukocyte activation; synapse formation;
KW transmembrane neurotransmitter receptor; autoimmune disease;
KW transplantation rejection; inflammation; allergy;
KW inflammatory bowel disease; ulcerative colitis; ileitis; psoriasis;
KW asthma; atopic dermatitis; atherosclerosis; cancer; infectious disease;
KW ischaemia; vasculitis; Crohn's disease.
XX
XX Homo sapiens.
OS
XX
XX WO200069898-A2.
PN
XX
XX 23-NOV-2000.
PD
XX
XX 12-MAY-2000; 2000WO-US13205.
PF
XX
XX 14-MAY-1999; 99US-0134114.
PR 14-MAY-1999; 99US-0134117.
XX 14-MAY-1999; 99US-0134118.
PR 21-OCT-1999; 99US-0160860.
XX 29-OCT-1999; 99US-0162498.
PR 13-DEC-1999; 99US-0170453.
XX 14-JAN-2000; 2000US-0176195.
PR 14-FEB-2000; 2000US-0182296.
XX 11-APR-2000; 2000US-0196460.
PR 11-APR-2000; 2000US-0196527.
XX
XX (ARBO-) ARBOR VITA CORP.
PA
XX
XX Lu PS;
PI
XX
XX WPI; 2001-061214/07.
DR
XX
XX Modulating a biological function of a hematopoietic cell for treating
PT an allergic response, or diseases mediated by immune system cells,
PT comprises introducing into the cell a PDZ-PL interaction enhancer or
PT inhibitor -
XX
XX
XX Disclosure; Page 38; 143pp; English.
PS
XX
XX The present invention relates to a method for modulating a biological
CC function of an endothelial cell or haematopoietic cell, comprises
CC introducing into a cell an antagonist that inhibits binding between a
CC PDZ domain protein and a PL domain protein to result in inhibition of
CC leukocyte activation. The present sequence is a PDZ domain. PDZ domains
CC of proteins are named after three prototypical proteins: PSD95,
CC Drosophila large disc protein and Zonula Occludin 1 protein. PDZ domain
CC proteins are involved in synapse formation by organising transmembrane
CC proteins are involved in synapse formation by organising transmembrane


```

XX 996 YRNRGRDLOSSTLSVPPQVMNHCSPSGSPHRYDVIGTRSMW---PSAPPGQRNVEQG 1052
DB 1012 -----elqge-----eeqllreerek-----rrrgeewgyrkkdelqgeeq 1050
QY 1053 HRGTRATGHTVITSRMDHRVMDHYSDRDRCADAROPHRHSRSTEQRLLETTTR 1112
DB 1051 lreer-----ekrrlqereryreeeelqgeeqllqgeetrirrelelryqrk 1099
QY 1113 SRSEPERDNTLMRSPSLMTGRSAPSPALSRSHPTGVSOTSPSPSTPGTGRGROLPOL 1172
DB 1100 eeeqlqgeeeqllre-----epkrrrgererqrceeeelqgeeeqllreerek 1147
QY 1173 PRKGLTESADIEERNRQMKLNKYQVAGSDPRLQDYHSKYRSGMDPH----- 1222
DB 1148 rrrgelelryreeeelqqrkryqr-----dedqrsdlkwqwepekenavrdnk 1198
QY 1223 --RGADTVSTKS--SDSDVSDVASVSRSTSSASRFSSTSYMSVQSERRGNRKISVFSKM 1279
DB 1199 yckgreneqfqrqledsqvrd-----rqsqdlqhlqgeeqdr-----ege 1240
QY 1280 NRQMGVSGKNLTKRSTISGDMCSLEKNDSOSDPAVGALGTSGKKRRSSIGAKMAIVGL 1339
DB 1241 rrrwqganrhfpeee-----qlereeqkea-----krr----- 1268
QY 1340 SRKRSASQLSQTEGGGKKRLRSTVQR 1365
DB 1269 drksgeekqllreereekrrrqrqtdr 1294

RESULT 11
AAB42491
ID AAB42491 standard; Protein; 3266 AA.
AC AAB42491;
DE 08-FEB-2001 (first entry)

Human ORFX ORF2255 polypeptide sequence SEQ ID NO:4510.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
vulnerary; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
immunosuppressant; thrombolytic; coagulant; vasotropic; antidiabetic;
hypotensive; dermatological; immunosuppressive; antinflammatory;
antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
neurodegenerative disorder; osteoarthritis; graft vs host disease;
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
cholesterol ester storage; systemic lupus erythematosus; infection;
severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
bone damage; cartilage damage; antinflammatory disease; coagulation;
thrombosis; contraceptive.

XX Homo sapiens.
OS
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000WO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinketsu RA, Iseach M;
XX
XX WPI; 2000-603262/57.
XX
XX N-PSDB; AAC76700.

```

```

XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
PS
XX
XX Claim 11; Page 3700-3708; 5507p; English.
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; coagulant; vasotropic;
CC immunosuppressant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; antiviral; antirheumatic;
CC antinflammatory; antibacterial; antifungal; antiproliferative;
CC antihypertensive; antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 3266 AA:
SQ
XX
XX Query Match 3.3%; Score 270.5; DB 21; Length 3266;
XX Best Local Similarity 20.6%; Pred. No. 1.7e-09;
XX Matches 316; Conservative 182; Mismatches 584; Indels 453; Gaps 68;

QY 268 DMPDRK-----RSPSVSDONRRYQSEEREDYQVPSDSDYMPSPSDYADRRSQREP 322
DB 222 dypqdrlyesvrltpqlypedsrtdypargrefyewekyqg-----dyesr----- 269
QY 323 QYEEPGHLYNRDSDNRGRHSKEYIVDEDEYERQREYERQREYERQREYERQREYERQREY 382
DB 270 -yyddp--reydyr-----ndpyeqdreyysrqr----- 299
QY 383 PVKQPYEOMRIHAESRAR-HERKHSVSIANAELERISLRLMDRPSRQSVSEER 441
DB 300 -----ererefesdrdrheerr-----piersgprvnlrrpgspaspsgerl 344
QY 442 AAMENQSYSMERTREAGQSSYQPTSNHSPPTPRRPIPLDRMDRADSRLKQHHL 501
DB 345 psdserrlys-----tsdsrsgscsisspryekldksrlerylknekt 389
QY 502 PSSAVRKTREKMETLRNDLSLSDQSESVRPPPRRHKSQKGMQVSLSSS-----EE 557
DB 390 kerrfdeperererrllrkevekdkt-----kqkrkykvnpspsgetden 439
QY 558 ELASTPEY-TCDDVLESESVSEKGDQSGK-----KRTSEQGVLSDSNTRSERQ 607
DB 440 ereqspekprscn--Kisrekadkeglaknrlmpcvvltrvekegkvaid-hpyekl 496
QY 608 KRMYYGSHLEDELEWSEPOIKDSGVDTCSSTLNEBHSKDPVYTWQSKGDRILG 667
DB 497 kakt-----dn-dtvksalidqklyvsqtepaksdisk----- 528
QY 668 RILNKLKDGSVPRDSGAMIGLVYGGKMTESGLCAFTIRKGGSLADYVGHLPD- 726
DB 529 --lesvrmk--vpkexylshvev-----ekgrl-----kark-----hlkpeqp 566
QY 727 -----EVLNNGRLLQGAFFE-----EYVNIILSKPEPOV---ELVY 761
DB 567 adgysavdlelearkrrfadsnlkaekqpevkkspsmedarvlskqpkvssrevll 626
QY 762 SRPIGDIPIRIDSTHAQLESSSSSFESQKMDR-----PSISVTSPMSPGMJR 808

```

```

Db      627 lre-geaerkp-----vrkeellkreekkikldrlntvaspdcqelasjsgsgrps--- 678
QY      809 DVPPQLSGQLSIKIMFKVGVQQLVITILGADLPBREDCGRPNPYK-lyFL-----PDSS 863
Db      679 ---sdlgarly-elagesvenq---evsgskkripisk-----pqkqqlvlddqpere 724
QY      864 DKNRRTKTYKVTLEPKKNQTFIYSP-----VHRREFRERLLETILMDQAVRE 913
Db      725 dvrykysalrdetpekrsgqgkshsvnteekigldihtqyrkqmgqgrrtkymemeia 784
QY      914 ESEFLG---ELLILELFTALL-----DDEPHWYKLTQTHDVSSLPRLPSPYLP 957
Db      785 ksekfsgpkkdvdeyerrslvhvqkppdvtrddspskkkrmhivdfdicckrennys 844
QY      958 RROUH-----GESPTRLQLOSKRISDSEVSDYCEDGVG-----VNSDYRHNGDLQS 1005
Db      845 srqtsedsergtgspsvr-----hgsfthededpigsprllsvkqspkvdekvlpy 894
QY      1006 STLAVPEQVMSNNHCSPSGSPHRVDVIGRTSRWSPPADPPQNVNGGHR-CTRATGHN 1063
Db      895 snltvreeslfnpydserrqmadmakiklsvlnsedelnwdsqmkqdaqrfdvfrpn 954
QY      1064 TISRMD--RRRVMD-----HYSRDRDCEADARQPYHRSRSTEQRLERTTSSRSE 1117
Db      955 slkrdsirkrsvrdlepgevpsdsdedge-----hksbprsalyle-----srls- 1002
QY      1118 RPDNLNMRSMNPLMTGRAPSPALSRHPRTGSVQTSPPGSGTGGRGRLQPLPDKST 1177
Db      1003 ---flirtdrecklrederlsslslern-kfysfalaktltpdt-----kal 1044
QY      1178 LER-----SAMDIERNRQMKLNRYKOYAGSDPR-LEDYHSKYSRGMQPHNG 1224
Db      1045 lerksjssstreenwstfldwdrfanfnmkxkdvdsapripipswymkkkirtsdsegk 1104
QY      1225 ADVTSTKSSDSVDVSAVSRSTSSAFSTSYMSVOSERPRGNKIKSYFTSKMONRQMG 1284
Db      1105 md-----dkkedhkeeeqegqelafsrflhsifegdekr-----lqlhkerkeesdt- 1152
QY      1285 VSGKNLTKSTISGDMCS-----LEKNDGSOS 1311
Db      1153 lsgriylykqtseganstdsldqepvylfhsrfmeltrmqgkkkdkqpkvekegedten 1212
QY      1312 DTAVGALGTSGK-----KRSSIGAKMVAIVGLSRKSRASOLSQTHGGGKKLRSYQBST 1367
Db      1213 hpkpesapenkdselktcpvsvgppsvtrvltl--esapsalektlg-----dktv 1260
QY      1368 ETGLAVENRMNMTROASRSTGDSNNYS-----SEGNLIFPGVRLAS 1410
Db      1261 eaplvteekltepatvseeakpasepapayeqleqvdlppqadpdkkaampayveegs 1320
QY      1411 DSQGSDFLDIGGPA-----QLVGRQTLATPAM-----GDIQVGMMDKKGOL 1451
Db      1321 sgddpypyladkprtpgasisfgaesnvdpbdpstqplskpqakeeaneppkdpdatada 1380
QY      1452 EVEL---IRAGGLVWPKSKTLPAVYKVVILLDNGVCLAKKTKYARKTLEFLYQOLLSF 1508
Db      1381 eptdnqkaeaaepesqpsedle-----vdppvaakdkkpnksktrskcpvgaavaai 1432
QY      1509 EESFOGRVLIQIIVWGDYGRMHK-----SFMGVAQIILL-ELLENNMYIGWFKLFFPSS 1561
Db      1433 vekpvtktse-----ridreklkrnsprgeaqkillelmeekilt-----rtaskns 1480
QY      1562 LVP-----PTSAPL--TRASQSSLESSTGPGYSRS 1590
Db      1481 aadlnehpepslipstlrtrnvrsyatmngdheuts 1515

```

```

XX      16-JAN-2001 (first entry)
DT
XX
XX      DE      Human APC protein (splice variant 2).
XX      APC gene; Adenomatous Polyposis Coli gene; human; chromosome 5q21;
XX      familial adenomatous polyposis; FAP locus; Gardner's syndrome; GS;
XX      sporadic tumour; adenoma; carcinoma; cancer; lung; breast; colon; rectum;
XX      bladder; liver; sarcoma; stomach; prostate; leukaemia; lymphoma;
XX      tumour suppressor; anti-APC antibody; detection; diagnosis; prognosis;
XX      genetic predisposition; drug screening; DP2.5; splice variant.
XX      Homo sapiens.
XX      US6114124-A.
XX      05-SEP-2000.
XX      25-MAY-1995; 95US-0450582.
XX      16-JAN-1991; 91GB-0000962.
XX      16-JAN-1991; 91GB-0000963.
XX      16-JAN-1991; 91GB-0000974.
XX      16-JAN-1991; 91GB-0000975.
XX      08-AUG-1991; 91US-0741940.
XX      12-AUG-1994; 94US-0289548.
XX      (ICIL ) IMPERIAL CHEM IND PLC.
XX      (UYJO ) UNIT JOHNS HOPKINS.
XX      (UTAH ) UNIT JOHNS HOPKINS.
XX      (CANC-) CANCER INST.
XX      Carlson M, Groden J, Joslyn G, Kinzler K, Markham AF, Anand R,
XX      Albertsen H, White RL, Thliveris A, Nakamura Y, Vogelstein B,
XX      Hedge PJ.
XX      WPI: 2000-565003/52.
XX      N-PSDB; AAB34350.
XX      Detecting Adenomatous Polyposis Coli (APC) protein in a sample for
XX      diagnosing cancers, involves contacting the sample with antibodies that
XX      specifically bind to APC protein and detecting the complex formed -
XX      Example 7; Fig 7A1-7W; 125pp; English.
PS
XX      The invention relates to a novel method for detecting Adenomatous
XX      Polyposis Coli (APC) protein in a sample. The method involves
XX      contacting the sample with antibodies which specifically binds to the
XX      2843 amino acid form of the human APC protein, or to a mutant APC
XX      protein, and detecting an APC-antibody complex. Mutations in the APC
XX      gene play a role in tumorigenesis, indicating that it is a tumour
XX      suppressor gene. It is located on chromosome 5q21, which corresponds to
XX      the FAP (familial adenomatous polyposis) locus. FAP is an autosomal
XX      dominant inherited disease in which affected individuals develop
XX      hundreds to thousands of adenomatous polyps in the colon and rectum.
XX      Some of which progress to malignancy. The FAP locus is often found to
XX      be deleted in sporadic (i.e., non-familial) adenomas and carcinomas, and
XX      chromosome 5q deletions have also been observed in tumours of the lung,
XX      breast, colon, rectum, bladder, liver, sarcomas, stomach, and prostate,
XX      and in leukemias and lymphomas. Although the FAP locus contains
XX      several other genes such as FER, TBI, TB2, and MCC, it is thought that
XX      mutations in the APC gene play a key role in the development of FAP and
XX      sporadic tumours. The method is useful for detecting APC protein and its
XX      mutant forms in foetal tissue, placental tissue, amniotic fluid, blood,
XX      serum or a tumour sample. The method is useful for diagnosing or
XX      prognosing neoplastic tissue, for detecting a genetic predisposition to
XX      cancer, for detecting germline and somatic alteration of wild-type APC
XX      genes, and for testing therapeutic agents for the ability to suppress
XX      tumours. The present sequence represents a 2742 amino acid splice
XX      variant of the human APC protein. This variant is less abundant than the
XX      2843 amino acid variant (AAB3011).
SQ      Sequence 2742 AA:

```



```

PR 16-JAN-1991: 91GB-0000963.
PR 16-JAN-1991: 91GB-0000974.
PR 16-JAN-1991: 91GB-0000975.
XX
PA (CANC-) CANCER INST.
PA (ICIL-) IMPERIAL CHEM IND.
PA (UYJO-) UNIV. JOHNS HOPKINS.
PA (UTAH-) UNIV UTAH.
XX
PI Albertsen H, Anand R, Carlson M, Groden J, Hedge PJ,
PI Joslyn G, Kinzler K, Markham AF, Nakamura Y, Thliveris A,
PI Vogelstein B, White RL,
XX
DR WPI: 1994-316233/39.
XX N-PSDB: AA072297.
XX
PT New human adenomatous polyposis coli DNA encoding tumour
PT repressor - and derived primers and probes for diagnosis,
PT prognosis and treatment of cancer
XX
PS Claim 3; Columns 71-84; 113pp; English.
XX
CC AA072297 is a cDNA isolated from the human adenomatous polyposis
CC coli (APC) gene, it encodes the tumour repressors described in AAR63507
CC and AAR63508. Determination of alterations in APC or its expression
CC products, can be used for the diagnosis and prognosis of cancer e.g.
CC colorectal, lung and breast tumours, and for determining predisposition
CC to certain cancers such as familial adenomatous polyposis (FAP) and
CC Gardner's syndrome. The wild type APC gene (or a part of it) can be
CC used therapeutically to restore gene function, while primers and probes
CC derived from the cDNA (AA072333-400 and AA072541-568) can be used to
CC detect mutations. Also APC proteins or analogues can be administered to
CC compensate for a defective gene, and epithelial cells, or transgenic
CC animals carrying a mutated APC allele are useful for detecting
CC therapeutic agents able to suppress tumorigenesis.
XX
SQ Sequence 2842 AA:

```

```

Db 1356 gakspsksgatpkspehyqetplmfrctsvs----sidsfesrlaavsgpcsg 1411
QY 460 -----GGSVPQRTSNHSPPTPRPRSPRLDRPDMRRADSLRKQNHLDSSAVRKTREK 513
Db 1412 mvsglispstlpsdpqgtmpsrsktppp-----ppqta--qtkre- 1450
QY 514 METPLRNDLSLDQSEGVPRPPPHKSKKCGKMRQVSLSSSEEL-----ASTPEYNSC 568
Db 1451 ---vprkpktaekres-----gpkgaavnaavgvqvlpdadlllfatesfpdgfsc 1501
QY 569 D-----DVEL-----ESESVEKGDGQ-KGKRKT--SEOG 595
Db 1502 sssaisaldepfiqkdvrlmpvpgdngnetesepkxeneqekaekitidsek 1561
QY 596 VLSDSN-----TSEKQKKMYGCGHSLFEDLEMSFQIKDSGVDTGSS 639
Db 1562 lldsdiddiellaeclisamptksrkkpqaasklppvarkpsql-----pyvk 1615
QY 640 TTLNEESHSDHPVTPQPSKDGRL--IGRLLN-----KRUKDSVPRDSGAMGL 690
Db 1616 llpsgnrlpqkh-vstfpgddmptvycvegtplnfstatslstdtsspnelaageyv 1674
QY 691 KVVGKMTESGRLCAFTTKYKGLADTVGHLRPQDEVLEWNGRLLOGATEEYVNIILE 750
Db 1675 r--gg--agsg-----efekrttlpgeg--rstdea-----ggktsavtllpeid 1713
QY 751 SKPEPQVELVY-----SRPID-----IPRLPDTHAQLESSESSFSQKMDRSISVT 799
Db 1714 dnkaeedillaeclinsampkgsnhkpfvkkimdvqg--qaasassaapkhngldgkkrpt 1772
QY 800 SPMSPGMLRDVPOLSQSLIKLMEFDKQGHQILVILGAKDLPSSEDDRPRPYVKIYFL 859
Db 1773 spvkr-----ipqnteyrlv-----tknadsknlnlaevf 1804
QY 860 PDRSDKKRRRTKTVKKTLEPK-----WNQFTYSPVH-----RR 893
Db 1805 sdnkdsqkqnlknskdnklpnedrvrgsfatdsphhytplegtrpycfersndsl 1864
QY 894 EFERMLETITMDARVKEESEFLGELILELTAALDDERPMWYLGQHDVSSLPR-- 951
Db 1865 dfddddvdlsr-ekaelrkaekenseekvsheltnsgsaanktga--lakgplnrgq 1921
QY 952 PSPYLPRLQ-----HGESPTRLQV-----SKRISDEVSVDYDCE----- 987
Db 1922 kpillgqstfpgsksktdptdgaatdeklnqfalentpvcfshmslslsldidennk 1981
QY 988 -----DGVGVVSDYRHHG-----RDLQSTLSVPRQVSSNHC 1020
Db 1982 enepiketepdsggepskpgasgyapksfhivedpvcfsrnsisslidseddllqec 2041
QY 1021 SPSGSPHRVD---VIGTRSWSPSAPPQRVNBOGHRKTRATGAYNTISKMD-RIRVMD 1076
Db 2042 issampkkrkpsrllkgnekhs-----rm-gqilgedlildkldqrpdsenglsp 2094
QY 1077 HYSRDDRDRCDAAD-----RO-----PYHRSRST 1101
Db 2095 sentdwaigegansivslngaaaaacistrgasdsdlsllksqisgspfhltpdge 2154
QY 1102 QRPLL-----ERTTTRSSRSERPDTN----- 1122
Db 2155 ekpftsmkgprllkpgektletkklieeskykgkkyvkslltgkvrnselsgmqk 2214
QY 1123 -LMSRMSLMTGRSAPSPALSRSHPRGSV-----QTSSTSTPGGRKGRQLPOLP 1174
Db 2215 plqanmpsisrgrtmlhpgvrnsssslspsvskkypplktlpaskspseg---qatatsp 2270
QY 1175 KGLTERSAMDIIEERNQKMLNKYKOVAGSDPRLLEDYISKRSRG---MDPHRGAP----- 1226
Db 2271 rgakpsvkselsparqt-----sqigss-----kapsrsgsdstpsrpqapqlsr 2318
QY 1227 TVSTKSSDSQSDVSAVSRSSASRFSSTYSVOSERPGRKRTISVFTSKMQRMGVS 1286
Db 2319 plqspgrnslspgrfnlspkllsqllptsspsstastksqsgkms-yts--pqrqm--s 2373

```



```
Db 1676 r--g--agsg-----efekrtipteg--rstdea-----ggktsvtilpeid 1714
OY 751 SKPEQVELV-----SRPIGD-----IPRIPOSTHAOLESSSSFPESQKMDRPSISVT 799
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1715 dnkaeegdllaecinsampkqshkpftrvkkimdvq--qasassapnknqldgkkkkpt 1773
OY 800 SPMSPGMLRDVPOPLSOLSIKLWFDKGHOLIYIILGAKDLPSREDDGRPRNPYKITYFL 859
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1774 spvkr-----ipqnteyletrv-----fkndasknlnaervf 1805
OY 860 PDRSDKNKRRTKYKKTLEPK-----WNQTFIYSPVH-----RR 893
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1806 sdnkdsckgnlknskdlnklnmedrvgsfafdsphhyplegtpcyfsrndslsl 1865
OY 894 EFRERMLEITLMDQARVEESEEFLGELLLEFALLDDEPHWYKLOTHDVSSLPLPR-- 951
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1866 dfddddvdlstr--ekaelrtkakenkesekvtshteltangqsanktqa--lakpinqrl 1922
OY 952 PSPYLRRLRQL-----HGESPTRLQR-----SKRISDSEVSDYDCE---- 987
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1923 pkpiliqyqstfpgsskdiptdgaadexlqnfalientpvcfshmslslsldidgenmk 1982
OY 988 -----DGVGVDYDRHNG-----RDLQSTLSVPEQVMSNHC 1020
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1983 enepiketepdpdgqepskpqasgyapksfhwedtpvcfsrnslslsldseddlqec 2042
OY 1021 SPSSGSPHREV--VIGTRMSPPSAPPPORNEYGHRGTRATGHYNTISRMD--RHRVMD 1076
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2043 issampkkkksrllkghnekhspr-----rnm-ggllgedlcldkldigrpdsbhglspd 2095
OY 1077 HYSDDRDRDCFEAD-----RQ-----PYHRSSTE 1101
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2096 sentdawkaiqgansivslhqaaaaaclsrqassdsdlsllksqslsgspflhlpdqe 2155
OY 1102 QRPL-----ERTTTRRSRSERPDTN----- 1122
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2156 ekpitsnkgprllkpgektletkkleesekyikgkkykalltgkvrnsneisgmky 2215
OY 1123 -LMRSMPLMTGRSAPSPALSHPRTGSV-----OTSPSSTPGTGRGRQLPOLPP 1174
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2216 plqanmpisrgrtmhlpvyrnsstspvskkypplktpraskspseg---qtaatsp 2271
OY 1175 KGTLERSAMDIEERRQKMLNKYQVAGSDPRLEDDYHSKYRSG--WDPHRCAD----- 1226
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2272 rgaqpsvkselsparqt-----sqiagss-----kapsrsgsrdsctprpaqplsr 2319
OY 1227 TVSTKSSDSDVDVSAVARTSASRFSSTSYMVSVOSEPRPGNKKISVFTSKMONROMGVS 1286
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2320 plqspgrnsisprngisppnklslqprtspstastkssgkms-yls--pgrqm--s 2374
OY 1287 GKMLTKSTISIGDMCGLKNDGSGSDTAVGALCTSGCKRRRSIGAKMVAIVGLSKRSRSA 1346
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2375 qgnltkqclglsknaasiprsesaaskylnqnmnngank-----velsmstsk 2423
OY 1347 SOLSTEGGGKKL-----RSTVQRSTETGLAVEMRNWMTROASREST----- 1388
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2424 ssgsesdrseprvlvrgstflkeapspllrklleesaafeslpsrpaaptlrsgaqlpv 2483
OY 1389 -----DGSMSNYS 1397
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2484 lpslpldmislsths 2498
```

Search completed: November 21, 2001, 16:05:29
Job time: 320 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2001, 16:06:19 ; Search time 16.56 Seconds
(without alignments) 3289.023 Million cell updates/sec

Title: US-09-617-099b-1
Perfect score: 8285
Sequence: 1 MSAPLPGPRGPRPAPASQP.....TRRASQSLESSNGPSYSRS 1590

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 338 | 4.1 | 1021 | 1 YPT7_CAEEL | P41885 caenorhabdi |
| 2 | 307 | 3.7 | 1898 | 1 TRHY_HUMAN | P07709 rattus norv |
| 3 | 262.5 | 3.2 | 684 | 1 RP3A_RAT | P06846 bos taurus |
| 4 | 261.5 | 3.2 | 704 | 1 RP3A_BOVIN | P20504 homo sapien |
| 5 | 258.5 | 3.1 | 2843 | 1 APC_HUMAN | P12270 homo sapien |
| 6 | 255.5 | 3.1 | 2349 | 1 TPR_HUMAN | P70478 rattus norv |
| 7 | 255 | 3.1 | 2842 | 1 APC_RAT | O61315 mus musculu |
| 8 | 254.5 | 3.1 | 2845 | 1 APC_MOUSE | O75376 homo sapien |
| 9 | 254 | 3.1 | 2468 | 1 MAPB_HUMAN | P30415 mus musculu |
| 10 | 253 | 3.1 | 2440 | 1 NCRL_HUMAN | P30414 homo sapien |
| 11 | 248.3 | 3.0 | 1453 | 1 NKCR_MOUSE | P15146 rattus norv |
| 12 | 244 | 2.9 | 1462 | 1 NKCR_HUMAN | P22793 ovies aries |
| 13 | 242 | 2.9 | 1861 | 1 MAP2_RAT | O61687 mus musculu |
| 14 | 241 | 2.9 | 1549 | 1 TRHY_SHEEP | P37709 oryctolagus |
| 15 | 239.5 | 2.9 | 2476 | 1 ATRX_MOUSE | O60974 mus musculu |
| 16 | 233.5 | 2.8 | 1407 | 1 TRHY_RABIT | P47708 mus musculu |
| 17 | 230.5 | 2.8 | 2453 | 1 NCRL_MOUSE | P48634 homo sapien |
| 18 | 224 | 2.7 | 606 | 1 BAT2_HUMAN | P20357 mus musculu |
| 19 | 223 | 2.7 | 2142 | 1 BAT2_MOUSE | P46100 homo sapien |
| 20 | 221.5 | 2.7 | 1828 | 1 MAP2_MOUSE | O02952 homo sapien |
| 21 | 221 | 2.7 | 2375 | 1 ATRX_HUMAN | P34926 rattus norv |
| 22 | 218 | 2.6 | 1781 | 1 AKAC_HUMAN | O13428 homo sapien |
| 23 | 217.5 | 2.6 | 2774 | 1 MAPA_RAT | P38904 saccharomyc |
| 24 | 216 | 2.6 | 1815 | 1 UN13_CAEEL | O94442 mus musculu |
| 25 | 214 | 2.6 | 1411 | 1 TCOF_HUMAN | O94442 mus musculu |
| 26 | 208.5 | 2.5 | 1395 | 1 SCF1_YEAST | P14873 homo sapien |
| 27 | 208 | 2.5 | 2472 | 1 NCRL_MOUSE | P21263 rattus norv |
| 28 | 207.5 | 2.5 | 1490 | 1 CRK7_HUMAN | P78559 homo sapien |
| 29 | 207.5 | 2.5 | 2464 | 1 MAPB_MOUSE | O10411 sciliosacch |
| 30 | 206 | 2.5 | 1805 | 1 NEST_RAT | O01484 homo sapien |
| 31 | 206 | 2.5 | 2805 | 1 MAPA_HUMAN | |
| 32 | 205 | 2.5 | 1957 | 1 YD86_SCHPO | |
| 33 | 205 | 2.5 | 3924 | 1 ANK2_HUMAN | |

| | | | | | |
|----|-------|-----|------|---------------|--------------------|
| 34 | 203.5 | 2.5 | 1676 | 1 APSA_EMENTI | O00083 emerice11a |
| 35 | 200 | 2.4 | 927 | 1 NED4_HUMAN | P46934 homo sapien |
| 36 | 199.5 | 2.4 | 568 | 1 INVO_RAT | P48998 rattus norv |
| 37 | 199.5 | 2.4 | 3038 | 1 TRIO_HUMAN | O75962 homo sapien |
| 38 | 198 | 2.4 | 1658 | 1 YME7_YEAST | O03651 saccharomyc |
| 39 | 198 | 2.4 | 2004 | 1 MOZ2_HUMAN | O92794 homo sapien |
| 40 | 197.5 | 2.4 | 3969 | 1 HRX_HUMAN | O03164 homo sapien |
| 41 | 197 | 2.4 | 2150 | 1 SDG3_CAEEL | P34706 caenorhabdi |
| 42 | 196.5 | 2.4 | 1189 | 1 XJH6_YEAST | P47035 saccharomyc |
| 43 | 196 | 2.4 | 388 | 1 SYT3_RAT | P40748 rattus norv |
| 44 | 195.5 | 2.4 | 3130 | 1 DPO2_HUMAN | O60673 homo sapien |
| 45 | 195 | 2.4 | 1075 | 1 NFC3_HUMAN | O12968 homo sapien |

ALIGNMENTS

| RESULT | ID | YP77_CAEEL | STANDARD | PRT | 1021 AA. |
|---|--|---|----------|------|----------|
| AC | P41885 | 01-NOV-1995 (Rel. 32, Created) | | | |
| DT | 01-NOV-1995 (Rel. 32, Last sequence update) | | | | |
| DT | 01-FEB-1996 (Rel. 33, Last annotation update) | | | | |
| DE | HYPOTHETICAL 111.7 KDA PROTEIN F37A4.7 IN CHROMOSOME III. | | | | |
| GN | F37A4.7 | | | | |
| OS | Caenorhabditis elegans. | | | | |
| OC | Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitoidea; | | | | |
| OC | Rhabdilitidae; Peloderiinae; Caenorhabditis. | | | | |
| OX | NCBI_TaxID=6239; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=BRISTOL N2; | | | | |
| RA | Fulton L., Waterston R.; | | | | |
| RL | Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases. | | | | |
| CC | -1- SIMILARITY: CONTAINS 2 C2 DOMAINS. | | | | |
| CC | -1- SIMILARITY: SOME, TO YEAST YNL087W. | | | | |
| CC | | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL Outstation - | | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | | |
| CC | or send an email to license@isb-sib.ch). | | | | |
| CC | | | | | |
| DR | EMBL: U00032; AAA50635.1; - | | | | |
| DR | HSSP: P21707; IRSY. | | | | |
| DR | Wormpep: F37A4.7; CE00705. | | | | |
| DR | InterPro: IPR000008; - | | | | |
| DR | InterPro: IPR001565; - | | | | |
| DR | Pfam: PF00168; C2; 2. | | | | |
| DR | PRINTS: PR00360; C2DOMAIN. | | | | |
| DR | PRINTS: PR00399; SYNAPTOFAGMN. | | | | |
| DR | PROSITE: PS00499; C2_DOMAIN_1; 2. | | | | |
| DR | PROSITE: PS50004; C2_DOMAIN_2; 2. | | | | |
| KW | Hypothetical Protein; Repeat. | | | | |
| FT | DOMAIN 758 847 C2 DOMAIN. | | | | |
| FT | DOMAIN 898 987 C2 DOMAIN. | | | | |
| SO | SEQUENCE 1021 AA; 111749 MW; 947C38B7C37BA462 CRC64; | | | | |
| | | | | | |
| Query Match | | | | | |
| Best local Similarity 20.9%; Pred. No. 5.1e-09; | | | | | |
| Matches 197; Conservative 139; Mismatches 342; Indels 266; Gaps 43; | | | | | |
| | | | | | |
| OY | 90 | KQNEKEPOKHLQOEPMYKVKM---GEESQOQOEK----- | 130 | GPAP | 130 |
| DB | 114 | RSPTNSAOTGSGITAAE--OEHTOKVLAKAESKSKQOIGKAVDRLEKRRRATGNGV | 171 | | 171 |
| OY | 131 | T-GGIGHKTKF---ADGCGHNSYCQTKFCAR-CGGR--VSLSKNV--MWCNCLCRK | 179 | | 179 |
| DB | 172 | THGLLCH-TEFGILLASVYAMCVDCRKYVCGORNGVETTDVNGTGTGKVEIVFLCKTCE | 230 | | 230 |

| QY | 180 | QOELL-TKSGAMFYNSGSSNTLQORDKQVPRGL--- | RHEAPQCKKKKILHMQFOGAPG- | 23 | |
|------------|---|--|---------------------------------------|----------|-----|
| | | | | | |
| Db | 231 | AREVLLKSGAMFKFKE-MPEFQRRDDLPYVRYVTTNGILPNASSA---- <td>ATPLSGTPGG</td> <td>285</td> | ATPLSGTPGG | 285 | |
| QY | 235 | -----DLSEVPAVE----- | KGRAHGLTRQDTINKGSGVKHQAIDMPSDRKS----- | 276 | |
| | | | | | |
| Db | 286 | AGPQPMFMPTSSQMTTPPKWASPGVCSNPGLOMNGP---- <td>TSPLNPGTRRMTNGHGLE</td> <td>341</td> | TSPLNPGTRRMTNGHGLE | 341 | |
| QY | 277 | -PVSYSRQNNRREYEDSEREDYSQYV---- <td>SDGTMPSBDYADRSQRQRYEEEPGHL</td> <td>331</td> | SDGTMPSBDYADRSQRQRYEEEPGHL | 331 | |
| Db | 342 | FPRSSSGSEFVSGVPRRALNNKTRPGSTATTSPAPPTSTTPPSRRE----- | A | 392 | |
| QY | 332 | NYRDSNRRGHRSKEYLVDEEDVSEPRDEYRORREEEYQARYRSDPMIARYPVKQPQYEE | 391 | | |
| Db | 393 | MEFYSHTTAHNAHRLKSTDDDDSSPESRSTRTS----- | PHSL----- | 434 | |
| QY | 392 | QMRTHAEVSPRAHRRHSDVSLANALEDLSRISLRMDRPSRQORSERRAAMENDRSYS | 451 | | |
| Db | 435 | -----ATPSSYAHDTCH-DTSLPADVTRSIDSGVYGDHNSPOOS----- | GLTCS | 478 | |
| QY | 452 | MEKTRERAGQSSYQYQRTSNHS-PTTPRRSPRLDPRMKRADYSIKOHNLDPSSAAKTK | 510 | | |
| Db | 479 | SSSLTPLOOQASH-----DHSGSGTPTRR----- | ISNPD----- | 511 | |
| QY | 511 | REKMETMLRNDLSDDQSESVPRPPRPKSKKGGKKRQVSLSSSEELASTPEYTSQD | 570 | | |
| Db | 512 | RV-----AQASAGSTSLVTPPPP----- | ISSKTSPPNCSSPLNV----- | 545 | |
| QY | 571 | VELESSEVSEKGDQSKQKRRKT-SBQGLVLSDSNTRSEBQKRRMYGHSLEEDLEWSEPD | 629 | | |
| Db | 546 | -----MEHKSASTASSGGRNRVGSAPRYLNNHHAHNNQ----- | NANDINKKLISQTSR | 596 | |
| QY | 630 | KDSGVDCSSSTTLNDEHSHSKHNVYTNQSKDGBRILGRILKKRLKDSGVPRRSGMLG | 689 | | |
| Db | 597 | AESPLAASSSPSSPDDDTQOKN-----RRKDDGVRANSJQLRTSLDVAAP----- | 643 | | |
| QY | 690 | LKVVGKMTSEGRICAFITKVKKKSLADTVGHLPRGDEVLEMNRLLOGATFEEVYNIL | 749 | | |
| Db | 644 | -----VARISKNN----- | GH----- | IV | 655 |
| QY | 750 | ESKREPOV--ELVVSRIQDIPRIPOSTH--AQLESSSSFESQKMDRPSISVTSWSP | 804 | | |
| Db | 656 | SSEPTSTTSNQNTHTSVPIPVAPVPEEEBEKATTAESTSEPGAVIEPLIDIDEMLEP | 715 | | |
| QY | 805 | -----GMRLQV----- | POPLSGQSLIKLMDKYGHOILYILGKDLPSREDG | 844 | |
| Db | 716 | KHASRRRGDVGVRGFCILCFRSKORSIGSTLTLLVTHSHADKKUKMKHLIRKNIKAKADSN | 775 | | |
| QY | 848 | RPRRPPVYVILPDRSDKNKRRRTYVKTTEPKKNNQFIYSPVHRRFERREMEITLMDQ | 907 | | |
| Db | 776 | GFSDPPYAFHLLPGTKATKILKSTIETKLTLPENNMELMSYGIATEDKKEKILKVATYLDK | 835 | | |
| QY | 908 | ARVRESESEPLGEILIEETALDLDEPRHWYKLQTHDVSLPLRP | 951 | | |
| Db | 836 | DRI--GSDPLGETRIALK-KLNNEMKKEFNILYE--SALPVPQ | 873 | | |
| RESULT 2 | | | | | |
| TRHY_HUMAN | TRHY_HUMAN | STANDARD; | PRT; | 1898 AA. | |
| AC | 007283; | | | | |
| DT | 01-OCT-1994 (Rel. 30, Created) | | | | |
| DT | 01-OCT-1994 (Rel. 30, Last sequence update) | | | | |
| DT | 01-OCT-2000 (Rel. 40, Last annotation update) | | | | |
| DE | TRICHOHYALIN. | | | | |
| GN | THH OR TRHY OR THL. | | | | |
| OS | Homo sapiens (Human). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | |
| OX | NCBI_TaxID=9606; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |

```

RX MEDLINE=93280194; PubMed=7685034;
RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,
RA Steinhert P.M.;
RT The structure of human trichohyalin. Potential multiple roles as a
RT functional EF-hand-like calcium-binding protein, a cornified cell
RT envelope precursor, and an intermediate filament-associated (cross-
RT linking) protein.
RL J. Biol. Chem. 268:12164-12176(1993).
RN [2]
RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=93315897; PubMed=7686953;
RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinhert P.M.;
RA "trichohyalin": a structural protein of hair, tongue, nail, and
RT epidermis.
RL J. Invest. Dermatol. 101:655-715(1993).
CC -I- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -I- SUBUNIT: MONOMER (PROBABLE).
CC -I- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC THE FLILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC -I- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC THE EPIDERMIS.
CC -I- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STANDED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.
CC -I- PTM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
CC -I- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC FAMILY.
CC -I- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L09190; AAA65582.1;
CC PIR: A45973; A45973.
CC HSSP: P02633; IBOC.
CC MIM: 190370;
CC InterPro: IPR001751;
CC DR InterPro: IPR002048;
CC Pfam: PF01023; S.100; 1.
CC Pfam: PF00036; eHand; 1.
CC PROSITE: PS00018; EF_HAND; 1.
CC DR PROSITE: PS00303; S100_CABP; 1.
CC Repeat; Calcium-binding.
CC DOMAIN 1 91
CC CA_BIND 22 33 SITE I (LOW AFFINITY) (POTENTIAL).
CC CA_BIND 62 73 SITE II (HIGH AFFINITY) (POTENTIAL).
CC DOMAIN 314 390 6 X 13 AA TANDEN REPEATS OF
CC R-R-E-O-E-E-E-R-R-E-O-O-L.
CC REPEAT 314 326 1-1 (APPROXIMATE).
CC REPEAT 327 339 1-2 (APPROXIMATE).
CC REPEAT 340 351 1-3 (APPROXIMATE).
CC REPEAT 352 364 1-4.

```

```
FT REPEAT 365 377 1-5.
FT REPEAT 378 390 1-6.
FT DOMAIN 391 444 9 x 6 AA TANDEM REPEATS OF R-R-E-Q-Q-L.
FT REPEAT 391 396 2-1.
FT REPEAT 397 402 2-2.
FT REPEAT 403 408 2-3.
FT REPEAT 409 414 2-4.
FT REPEAT 415 420 2-5.
FT REPEAT 421 426 2-6.
FT REPEAT 427 432 2-7.
FT REPEAT 433 438 2-8.
FT REPEAT 439 444 2-9.
FT DOMAIN 444 702 9 x 28 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 923 952 8 x 30 AA TANDEM REPEATS.
FT REPEAT 923 952 4-1.
FT REPEAT 953 982 4-2.
FT REPEAT 983 1012 4-3.
FT REPEAT 1013 1042 4-4.
FT REPEAT 1043 1072 4-5.
FT REPEAT 1073 1102 4-6.
FT REPEAT 1103 1132 4-7.
FT REPEAT 1133 1162 4-8.
FT DOMAIN 1250 1849 23 x 26 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 1752 1752 F -> L (IN REF. 2).
FT CONFLICT 1794 1801 OERDROYR -> RSETGSGTG (IN REF. 2).
FT CONFLICT 1857 1857 O -> K (IN REF. 2).
FT CONFLICT 1880 1880 V -> G (IN REF. 2).
SQ SEQUENCE 1898 AA; 247219 MW; A7AB5947FB2E31D CRC64;
```

Query Match 3.7%; Score 307; DB 1; Length 1898;

Best Local Similarity 17.4%; Pred. No. 3.3e-07; Matches 245; Conservative 213; Mismatches 514; Indels 434; Gaps 53;

```
QY 19 QPPPEPMDLSHLEEE--RKTILAVMDROKKEEKEQSVLKIEKHAOPTOWPFPFG 76
Db 264 EPPORRELOE-----EEBQLRKLEROELRLREROEEOOQRLRLRQOE----- 312
QY 77 ITELAVNVLOQOKOPNKKEPOTKLIHQFEMKKEOVKKKGESOOQ---EOKGAPTCG 133
Db 313 -----EREOOEEREOOEEREOOEEREOOEEREOOEEREOOEEREOOEEREOOE 358
QY 134 ICHRTFADGCGHNSYCQTKFCARGCGVSLRSNKVMVMCMCRKOELITKSGAMFYN 193
Db 359 -----REOQL----- 364
QY 194 SGSNTLOPDKVPRLGRNEAPQEKAKLHEOPQFQAGDLVPAVEKGRAGHGLTOD 253
Db 365 ---RREOEERREOQLRREOEERREOQLRREOQLRREOQLRREOQLRREOQLRREO 412
QY 254 TIKNGSGVGHOLASMPDRKSPSVSRDQNRRYFOSEER-EDYSQYVSDGTMPSFSD 312
Db 413 QLRREOQLRREOQLRREOQLRREOQLRREOQLRREOQEERHOKHEOEEREOQLR 472
QY 313 YAD-----RRSQREPOFYEEPGH---LNYRDSNRRGRHSKEYITVDEDEVE 355
Db 473 RRDMLKREETERHNEQERKQQLKRDQEEERERMLKLEEEERREOQERREOQLRREOE 532
QY 356 SRDE-YERORREEEYOAARYSDPNLARYVPVKOPYEOMRIHAEVSRAHERHSDVSLA 414
Db 533 RREORLKQEEERLOQLRREOQLRREOQLRREOQLRREOQLRREOQLRREOQLR 586
QY 415 NAELEDSTISLMDRPRQRQSVSRBRAMENQRSYSMKTREAGQSSYPORTSHSPR 474
Db 587 KREOEERDQQLKREERROQLKREOEERLEOR-LKREVERLEOEERDELEKREPR 645
QY 475 TPRSPTPLDRPDMRRASLKKOHNLDPSSAVARKTKREKMETMLRDSLSDQSERVPP 534
Db 646 EERRH-----ELKSEOEERRH-----EQLRREOQERREOQLRKE-----EEFERLEOR 650
QY 535 PPRPKSKGKGMROVSLSSSEELA-----STPEYTSQDVELESESVSEKGDQOKG 588
Db 691 LKREHEER-----RQELALEEOQARERIKSRIPKW-----OMOLESEA-----DARQSK 737
```

```
QY 589 -RKTSEGVLSDSNTRSEROKRMUYGSHLEEDLEWSEPOIKDSGVDTCSSTLNEHS 647
Db 738 VLEAPQAGRAEAPQEOEERKR-----ESELQWE-----ERRA 772
QY 648 H-----SDKHVYWPQSKDQDLRLILNKKRLKDSVPRDSGMLGLKVVGKMTESG 701
Db 773 HROQOEERQDRFTWQAOAEKSESRGQRLSAR-----PRLREQREQLRAEERQQRQR 827
QY 702 RLCAFTKKVKKGLADTVGHILRGDEVLEWNGRLLOGATFEEVYNIILIESRPEOVELV 761
Db 828 FL-----PEEKEQERGR-----QREREKELOF 851
QY 762 SRPIGDIRIPDSTHAQLESSSSFEQOKMDRPSISVTSFPMSPGMLRVPOFLSGLSIK 821
Db 852 -----LEF-----EOLQRRERAQLOEBEDGLQEOERROQERQORQD 889
QY 822 LMFQVGHOLIVTLGAKDLPSREDGRPRNPVVIYFLPDRSDKKARRTKTVKKTLEPKW 881
Db 890 KW-----RMQL-----EEKRRRRHT-LYAKPALQEOQLRQOQLQEEEEELO 931
QY 882 NQFTIYSPVHREFRPMLETTLMDQARVEESEFLCELLILETALLDDEPHWY----- 937
Db 932 REE---REKRROEOEROYR-----EEOLOQEEOLRLE---EREKRRROEREROYRDK 981
QY 938 KLQTHDVSSLPLRPSYLYPRROLHGESP--TRRLQREKRISSDSVSDYDCEDGVVSD 995
Db 982 KLOOKE-----EOLLGEPEKRRROERREKTYREE----- 1011
QY 996 YRHNGRDLQSTLTSVPQVMSNHCSPGSPHRYDVIGRTSWS---PSAPPOQRNVEG 1052
Db 1012 -----BLQE-----EQQLREERK-----RRQEWEROYKKKDELQOELO 1050
QY 1053 HRGTRATGHVNTISRMDHRYVMDHYSSDRDCEADRDQYHRSRSTEQRLPRTTTR 1112
Db 1051 LREER-----EKRLQEOEROYREEEELOQEELOLGEERETRRQOLEEROYRK 1099
QY 1113 SRSEERDYNLMRSMPLMTGCRSAPSPALSRSHPRIGSVYQTSPTSGTGRGQLOL 1172
Db 1100 EEBLOQEEBOLR-----BEKRRROERROEEOEBELOEEOLEERK 1147
QY 1173 PPKQTLERSAMDIEERNOMKLNKYQVAGSDPRLQDYHNSKYSGMPIH----- 1222
Db 1148 RRQLELROYREEBELOKQKQRYR-----DEQQRDQLKWQPEKENAVRDKV 1198
QY 1223 --RCADTVSTKS--SDSDVSDYSAVSRTSSASRFSSTYSVQSERPRGNKISVFTSKM 1279
Db 1199 YCKGRENEOFQLEDQSVRD-----ROSQODLOHLGEOERDR-----EOE 1240
QY 1280 NROKVSAGKMLTKSTJSTSGDMCSLEKNDGSGSDPAVAGLAGTSGKKRRSSIGAKMVAI 1339
Db 1241 RRRWQANRHPREE-----OLEREOKEA-----KRR----- 1268
QY 1340 SRKSRSASOLQTEGGGKKLSTVOR 1365
Db 1269 DRKSQEEKOLLREERREKRRRQETDR 1294
```

RESULT 3
R3A_RAT
ID R3A_RAT STANDARD; PRT; 684 AA.
AC P4709;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE RABPHILIN-3A.
GN RPH3A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A.

```

RX MEDLINE-95033210; PubMed-7946335;
RA Li C., Takei K., Geppert M., Daniel L., Stenius K., Chapman E.R.,
RA Jahn R., de Camilli P., Suedhof T.C.;
RT "Synaptic targeting of rabphilin-3A, a synaptic vesicle
RT Ca2+/phospholipid-binding protein, depends on rab3A/3C.";
RT Neuron 13:885-898(1994).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 44-167 IN COMPLEX WITH RAB3A.
RC TISSUE=Brain;
RX MEDLINE-99148269; PubMed-10025402;
RA Ostermeier C., Brunger A.T.;
RT "Structural basis of Rab effector specificity: crystal structure of
RT the small G protein Rab3a complexed with the effector domain of
RT rabphilin-3A.";
RT Cell 96:363-374(1999).
RL -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED WITH RAS-RELATED
CC PROTEIN RAB-3A IN SYNAPTIC VESICLE TRAFFIC AND/OR SYNAPTIC
CC VESICLE FUSION. COULD PLAY A ROLE IN NEUROTRANSMITTER RELEASE BY
CC REGULATING MEMBRANE FLOW IN THE NERVE TERMINAL.
CC -1- SUBUNIT: MONOMER. SPECIFICALLY EXPRESSED IN BRAIN.
CC -1- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN BRAIN.
CC -1- SIMILARITY: CONTAINS 2 C2 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U12571; AAA62662.1; -
CC PDB: 1ZBD; 12-APR-99.
DR InterPro: IPR000008; -
DR InterPro: IPR001565; -
DR InterPro: IPR001965; -
DR Pfam: PF00168; C2; 2.
DR Pfam: PF00628; C2; 2.
DR PRINTS: PR00360; C2DOMAIN.
DR PRINTS: PR00399; SYNAPTOTAGM.
DR PROSITE: PS00499; C2_DOMAIN_1; 2.
DR PROSITE: PS50004; C2_DOMAIN_2; 2.
DR Repeat: Synapse; protein transport; 3D-structure.
KW DOMAIN 397 645 PHOSPHOLIPID BINDING (PROBABLE).
FT DOMAIN 397 488 C2 DOMAIN.
FT DOMAIN 537 645 C2 DOMAIN.
SQ SEQUENCE 684 AA; 75832 MW; 05838BC3C7A86444 CRC64;

```

```

QY 332 NY-----KDSNRGRHSKEYIVDEDEVSREDEYERQ--RREEXQARY 373
DQ :::::
DB 225 SHGPPTRRASAEARMSTTTPDSGWDHGHG-----GGAGPTRSRPGEGGLRFRANSVOAS- 278
QY 374 RSDPNLIARYPVKRPQ-----YEEDMRTHAEYSRAHREHSHVSIANA 416
DB 279 RPAAPASMPSPAPQPPQPPGPGSRAAPGPGREPPEST-----EAPSPDPGPGA 328
QY 417 ELEDSTISILNRDPRQSVSRRAAMENQRSYMERPEAOGQSSYPPTKSNHSPPTP 476
DB 329 V-----APARE-----ERIGPTGFGQAAPPTIACPIGQAAP 358
QY 477 RSPPIIDPRDMKRADSLKROHLLDPSAVRKTRKREKMETMLRNDLSDD-QGSEVRRPP 535
DB 359 AKQPPAAEE-----EANSYDSQA--TTLGALFSLIYDDQNSMLQCTIIRAKG 408
QY 536 PRPHSKKKGKKRQVLSLSSEELASTPEYTSDDIVLESESYEKDQSGKGRKTEGQ 595
DB 409 LKPMDSN-----GLADPYVKLHLLP-----GASKSNKLRTK-- 439
QY 596 VLSDSMTREPROKKRMYYGSHLSLEEDLEWSEPOIKDSGVPTCSSTYLNEHSHDKHPVT 655
DB 440 --TLRNTRPVWNETLYQHCIT--EDDK--QRKTIKISVCD--EDKFGHNE----- 482
QY 656 WQPSKXDRILGRILLNKLKDGSVPRDSGAMGLKAVGKMTESGRICAPITKVKKGS 715
DB 483 -----FIGE-----TRFSLKKLKANQRKFNIC--LERVIDPKR 514
QY 716 ADIVGHLRDEDEVLWNGRLLOGATPEEYVNIILLESPEQVELVYRPIGDIRIPDST 775
DB 515 AGTTGSGAR-----GMAIYE-----EDQVE-----RIGDIE----- 540
QY 776 HAQLESSSSSFESQKMDRPSISVTPSPMSGMLRDVPOFLSGOLSIKLMKVGQHLIVTI 835
DB 541 -----GKILVSLMYSTQCGGLIYGI 561
QY 836 LGAKDLPSREDGRPRNDYKIYFLPDRSDKNKRTKTKVTKLEPKNMOTPIYSVHRRF 895
DB 562 IRCVHLAAMDANGYSDPFLWMLKPDWKKRKAHKTOIKTKTLIDNEFEPEYDIKH--SDL 620
QY 896 REMMLETTIMDOARVRESESPFGEILIELEVALLDDEPRHWK 938
DB 621 AKRSLDISVWDYD--IGKSDNYIGGCOLGI--SAKGERLKHMYE 660

```

RESULT 4

| ID | RP3A_BOVIN | STANDARD; | PRT; | 704 AA. |
|----|---|-----------|------|---------|
| AC | 006846; | | | |
| DT | 01-FEB-1996 (Rel. 33, Created) | | | |
| DT | 01-FEB-1996 (Rel. 33, Last sequence update) | | | |
| DT | 01-NOV-1997 (Rel. 35, Last annotation update) | | | |
| DE | RABPHILIN-3A. | | | |
| GN | RPH3A. | | | |
| OS | Bos taurus (Bovine). | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; | | | |
| OX | Bovidae; Bovine; Bos. | | | |
| OX | NCBI_TaxID=9913; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A., AND PARTIAL SEQUENCE. | | | |
| RC | TISSUE=Brain; | | | |
| RX | MEDLINE-93204952; PubMed-8384302; | | | |
| RA | Shirataki H., Kaibuchi K., Sakoda T., Kishida S., Yamaguchi T., | | | |
| RA | Wada K., Miyazaki M., Takai Y.; | | | |
| RT | "Rabphilin-3A, a putative target protein for smg p25a/rab3A p25 small | | | |
| RT | GTP-binding protein related to synaptotagmin."; | | | |
| RL | Mol. Cell. Biol. 13:2061-2068(1993). | | | |
| RN | [2] | | | |
| RP | DOMAINS. | | | |
| RX | MEDLINE-94086530; PubMed-8262955; | | | |
| RA | Yamaguchi T., Shirataki H., Kishida S., Miyazaki M., Nishikawa J., | | | |
| RA | Wada K., Numata S.-I., Kaibuchi K., Takai Y.; | | | |

```

RT      "Two functionally different domains of rabphilin-3A, Rab3a p25/smg
RT      p25a-binding and phospholipid- and Ca(2+)-binding domains."
RT      J. Biol. Chem. 268:27164-27170(1993).
CC      -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED WITH RAS-RELATED
CC      PROTEIN RAB-3A IN SYNAPTIC VESICLE TRAFFIC AND/OR SYNAPTIC
CC      VESICLE FUSION. COULD PLAY A ROLE IN NEUROTRANSMITTER RELEASE BY
CC      REGULATING MEMBRANE FLOW IN THE NERVE TERMINAL.
CC      -1- SUBUNIT: MONOMER.
CC      -1- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN BRAIN.
CC      -1- SIMILARITY: CONTAINS 2 C2 DOMAINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: D13613; BAA02780.1; -.
DR      HSSP: P21707; IRSY.
DR      InterPro: IPR000008; -.
DR      InterPro: IPR001565; -.
DR      InterPro: IPR001965; -.
DR      Pfam: PF00168; C2; 2.
DR      Pfam: PF00628; PHD; 1.
DR      PRINTS: PR00360; C2DOMAIN.
DR      PRINTS: PR00399; SYNAPTOTAGMN.
DR      PROSITE: PS00499; C2_DOMAIN_1; 2.
DR      PROSITE: PS50004; C2_DOMAIN_2; 2.
KW      Repeat: Synapse; Protein transport.
FT      DOMAIN 1 280 RAB P25/SMG P25A BINDING (GTP GAMMA-S-
FT      DOMAIN 281 704 PHOSPHOLIPID BINDING; CA2+-DEPENDENT.
FT      DOMAIN 417 508 C2 DOMAIN:
FT      DOMAIN 577 665 C2 DOMAIN.
SQ      SEQUENCE 704 AA; 77977 MW; 132AD048F58F8FED4 CRC64;

```

```

DB      297 -----GQPGPGGSRSPSPGTGR---FPDQRPPEVAPSDP-----DYTGAAAPREERTG 342
QY      516 TMLRNSLSDDQSESVRP-----PPRPHKSKGGMKQVYSLSS 555
DB      343 GIGGYAAGTREDRAGPPPSYSTQASAAADQPVYASARQPPPEDEBEAN-----SYSD 398
QY      556 EEELASTPEYT-----SCDDVESESEYSEKG---DSO-----KGKRTSEGV 596
DB      399 EATTLALERSLLYDODNSLHCTIIKAKGLPMDNSGLADPYKHLHLPASNSKRT 458
QY      597 LSDSNTRSEROKRMVYGGHSLDELDLEWSEBQIKDSDVTCSTTLNEHSHDKHPVTW 656
DB      459 KTLRNTRNPIWNETIYVHGIT-DEDM-----QRKTLRISVD-----EDKFGHNE----- 502
QY      657 QPSKDDRLIGRLLNKRLLDGSVPRDSGMLGLKVVGGKMTSEGRICAFITKXKSLA 716
DB      503 -----FTGE-----TRESLKRLKPNOKRKNFNC--LEVIYPMKRA 535
QY      717 DTVGHLPDGEVLEWNGRLLOGATFEFVYNILLESKEPQVELVYSRPIGDIPIRDPSTH 776
DB      536 GTTGSAR-----GMAIYE-----EEGYE-----RIGDIE----- 560
QY      777 AQLESSSSSFESQKMDRPSISVTSMPMSGLRDVQPLSGQLSIKLPMDKYGHOLIYTL 836
DB      561 -----RGLIVSIMYSTQGGGLIVGII 582
QY      837 GAKDLPSEDEGRPNRPVYKIYFLPDRSDKNKRTKYKTLTEPKMNQFITSVPHRRFR 896
DB      583 RCYHLAMDANGYSDPVPYKLMKLPDMGKRAKHKTQIKRLTLPENEEFEFYDIKH-SDLA 641
QY      897 ERMLEITLMDQARVYEESEFELEILELETALDDEPHWYK 938
DB      642 KKSLLDISWYD--IGKSNDYIGGCGQLGI-SAKGERLKHWIE 680

RESULT 5
APC_HUMAN STANDARD; PRT; 2843 AA.
ID APC_HUMAN
AC P25054; Q15162; Q15163;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ADENOMATOUS POLYPOSIS COLI PROTEIN (APC PROTEIN).
GN APC OR DP2.5.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91335210; PubMed=1651562;
RA Kinzler K.W., Nizard M.C., Su L.-K., Vogelstein B., Bryan T.M.,
RA Levy D.B., Smith K.J., Preisinger A.C., Hedge P., McKechnie D.,
RA Flumlear R., Markham A., Groffen J., Boguski M.S., Altshul S.F.,
RA Horii A., Ando H., Miyoshi Y., Miki Y., Nishisho I., Nakamura Y.;
RT "Identification of FAP locus genes from chromosome 5q21.";
RL Science 253:661-665(1991).
RN [2]
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RC TISSUE=Fetal brain;
RX MEDLINE=91330307; PubMed=1678319;
RA Joslyn G., Carlson M., Thiliveris A., Albertsen H., Gelbert L.,
RA Samowitz W., Groden J., Stevens J., Spirio L., Robertson M.,
RA Sargent L., Kravcho K., Wolff E., Burt R., Hughes J.P.,
RA Warrington J., McPherson J.D., Wasmuth J., Le Paslier D.,
RA Abderhagim H., Cohen D., Leppert M., White R.;
RT "Identification of deletion mutations and three new genes at the
RT familial polyposis locus.";
RL Cell 66:601-613(1991).
RN [3]
RP ASSOCIATION WITH CATENINS.
RX MEDLINE=94082295; PubMed=8259519;
RA Su L.-K., Vogelstein B., Kinzler K.W.;

```

RT "Association of the APC tumor suppressor protein with catenins.";
 RL Science 262:1734-1737(1993).
 RN [14]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-94154728; PubMed-8111410;
 RA Nagase H., Nakamura Y.;
 RT "Mutations of the APC (adenomatous polyposis coli) gene.";
 RL Hum. Mutat. 2:425-434(1993).
 RN [15]
 RP VARIANTS FAP.
 RX MEDLINE-91335211; PubMed-1651563;
 RA Nishisho I., Nakamura Y., Miyoshi Y., Miki Y., Ando H., Horii A.,
 RA Koyama K., Utsunomiya J., Baba S., Hedge P., Markham A., Krush A.J.,
 RA Petersen G., Hamilton S.R., Nilbert M.C., Levy D.B., Bryan T.M.,
 RA Peisinger A.C., Smith K.J., Su L.-K., Kinzler K.W., Vogelstein B.;
 RT "Mutations of chromosome 5q21 genes in FAP and colorectal cancer
 RT patients.";
 RL Science 253:665-669(1991).
 RN [16]
 RP VARIANTS FAP.
 RX MEDLINE-93265030; PubMed-1338904;
 RA Miyoshi Y., Nagase H., Ando H., Ichii S., Nakatsuru S., Aoki T.,
 RA Miki Y., Mori T., Nakamura Y.;
 RT "Somatic mutations of the APC gene in colorectal tumors: mutation
 RT cluster region in the APC gene.";
 RL Hum. Mol. Genet. 1:229-233(1992).
 RN [17]
 RP VARIANTS FAP.
 RX MEDLINE-93244793; PubMed-1338691;
 RA Nakatsuru S., Yanagisawa A., Ichii S., Tahara E., Kato Y.,
 RA Nakamura Y., Horii A.;
 RT "Somatic mutation of the APC gene in gastric cancer: frequent
 RT mutations in very well differentiated adenocarcinoma and signet-ring
 RT cell carcinoma.";
 RL Hum. Mol. Genet. 1:559-563(1992).
 RN [18]
 RP VARIANTS FAP W-1348, AND VARIANTS D-1118; M-1292; V-1304 & S-2502.
 RX MEDLINE-93250848; PubMed-1338764;
 RA Nagase H., Miyoshi Y., Horii A., Aoki T., Petersen G.M.,
 RA Vogelstein B., Maher E., Ogawa M., Maruyama M., Utsunomiya J.,
 RA Baba S., Nakamura Y.;
 RT "Screening for germ-line mutations in familial adenomatous polyposis
 RT patients: 61 new patients and a summary of 150 unrelated patients.";
 RL Hum. Mutat. 1:467-473(1992).
 RN [19]
 RP VARIANTS FAP TRP-99.
 RC TISSUE-Peripheral blood lymphocytes;
 RX MEDLINE-95134544; PubMed-7833149;
 RA Dobbie Z., Spycher M., Huerliman R., Ammann R., Ammann T., Roth J.,
 RA Mueller A., Mueller H., Scott R.J.;
 RT "Mutational analysis of the first 14 exons of the adenomatous
 RT polyposis coli (APC) gene.";
 RL Eur. J. Cancer 30A:1709-1713(1994).
 RN [10]
 RP VARIANTS FAP GLY-722.
 RX MEDLINE-95135430; PubMed-7833931;
 RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,
 RA Romio L., Pilla S., Prete F., Mareni C., Guanti G.;
 RT "Four novel mutations of the APC (adenomatous polyposis coli) gene in
 RT FAP patients.";
 RL Hum. Mol. Genet. 3:1687-1688(1994).
 RN [11]
 RP ERRATUM.
 RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,
 RA Romio L., Pilla S., Prete F., Mareni C., Guanti G.;
 RL Hum. Mol. Genet. 3:1918-1918(1994).
 RN [12]
 RP VARIANTS FAP ILE-171.
 RX MEDLINE-97144176; PubMed-8990002;
 RA van der Luijt R.B., Meera Khan P., Vasen H.F.A., Tops C.M.J.,
 RA van Leeuwen-Cornelisse I.S.J., Wijnen J.T., van der Klift H.M.,
 RA Plug R.J., Griffioen G., Fodde R.;
 RT "Molecular analysis of the APC gene in 105 Dutch kindreds with

RT familial adenomatous polyposis: 67 germline mutations identified by
 RT DGGE, PCR, and southern analysis.";
 RL Hum. Mutat. 9:7-16(1997).
 RN [13]
 RP VARIANTS LYS-1307.
 RX MEDLINE-98400248; PubMed-9731522;
 RA Redston M., Nathanson K.L., Yuan Z.O., Neuhausen S.L., Satagopan J.,
 RA Wong N., Yang D., Nafa D., Abrahamson J., Ozcelik H.,
 RA Antin-Ozertis D., Andrulis I., Daly M., Plinsky L., Schrag D.,
 RA Gallinger S., Kaback M., King M.-C., Woodage T., Brody L.C.,
 RA Godwin A., Warner E., Weber B., Foulkes W., Offit K.;
 RT "The APC I1307K allele and breast cancer risk.";
 RL Nat. Genet. 20:13-14(1998).
 RN [14]
 RP VARIANTS LYS-1307 AND GLN-1317.
 RC TISSUE-Peripheral blood;
 RX MEDLINE-98393712; PubMed-9724771;
 RA Frayling I.M., Beck N.E., Ilyas M., Dove-Edwin I., Goodman P.,
 RA Pack K., Bell J.A., Williams C.B., Hodgson S.V., Thomas H.J.W.,
 RA Talbot I.C., Bodmer W.F., Tomlinson I.P.M.;
 RT "The APC variants I1307K and E1317Q are associated with colorectal
 RT tumors, but not always with a family history.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:10722-10727(1998).
 RN [15]
 RP VARIANTS LYS-1307.
 RX MEDLINE-98400259; PubMed-9731533;
 RA Woodage T., King S.M., Wacholder S., Hartge P., Struwing J.P.,
 RA McAdams M., Laken S.J., Tucker M.A., Brody L.C.;
 RT "The APC I1307K allele and cancer risk in a community-based study of
 RT Ashkenazi Jews.";
 RL Nat. Genet. 20:62-65(1998).
 RN [16]
 RP VARIANTS LYS-1307.
 RX MEDLINE-99138651; PubMed-9973276;
 RA Gryfe R., Di Nicola N., Lai G., Gallinger S., Redston M.;
 RT "Inherited colorectal polyposis and cancer risk of the APC I1307K
 RT polymorphism.";
 RL Am. J. Hum. Genet. 64:378-384(1999).
 RN [17]
 RP VARIANTS FAP CYS-1171 & THR-2738, AND VARIANTS GLY-1057 & VAL-1822.
 RX MEDLINE-9913859; PubMed-9950360;
 RA Wallis Y.L., Morton D.G., McKeown C.M., MacDonald F.;
 RT "Molecular analysis of the APC gene in 205 families: extended
 RT genotype-phenotype correlations in FAP and evidence for the role of
 RT APC amino acid changes in colorectal cancer predisposition.";
 RL J. Med. Genet. 36:14-20(1999).
 RN [18]
 RP VARIANTS FAP PRO-1184.
 RX MEDLINE-99401091; PubMed-10470088;
 RA Iamlium H., Ilyas M., Rowan A., Clark S., Johnson V., Bell J.A.,
 RA Frayling I.M., Efsthathiou J., Pack K., Payne S., Roylance R.,
 RA Gorman P., Sheer D., Neale K., Phillips R., Talbot I.C., Bodmer W.F.,
 RA Tomlinson I.P.M.;
 RT "The type of somatic mutation at APC in familial adenomatous polyposis
 RT is determined by the site of the germline mutation: a new facet to
 RT Knudson's 'two-hit' hypothesis.";
 RL Nat. Med. 5:1071-1075(1999).
 RN [19]
 RP FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-
 CC CATEININ. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE
 CC ALLOWS THE DOWNREGULATION OF CYTOPLASMIC BETA-CATEININ.
 CC SUBUNIT: FORMS HOMOLOGOMERS AND ASSOCIATES WITH CATENINS.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM. ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TISSUES TYPES.
 CC -1- PTM: PHOSPHORYLATED BY GSK-3B.
 CC DISEASE: DEFECTS IN APC ARE A CAUSE OF FAMILIAL ADENOMATOUS
 CC POLYPOSIS (FAP) AND GARDNER'S SYNDROME (GS), THAT CONTRIBUTE TO
 CC TUMOR DEVELOPMENT IN PATIENTS WITH NONINHERITED FORMS OF
 CC COLORECTAL CANCER. FAP IS CHARACTERIZED BY ADENOMATOUS POLYPS OF
 CC THE COLON AND RECTUM, BUT ALSO OF UPPER GASTROINTESTINAL TRACT
 CC (AMPULLARY, DUODENAL, AND GASTRIC ADENOMAS). THIS IS A VICIOUSLY
 CC PREMALIGNANT DISEASE WITH ONE OR MORE POLYPS PROGRESSING THROUGH
 CC DYSPLASIA TO MALIGNANCY IN UNTREATED GENE CARRIERS WITH A MEDIAN

```

CC      AGE AT DIAGNOSIS OF 40 YEARS.
CC      DISEASE: APC. MUTATIONS HAVE LED TO SOME INTERESTING OBSERVATIONS.
CC      (1) THE GREAT MAJORITY OF THE MUTATIONS FOUND TO DATE WOULD RESULT
CC      IN TRUNCATION OF THE APC PRODUCT. (2) ALMOST ALL THE MUTATIONS
CC      HAVE OCCURRED WITHIN THE FIRST HALF OF THE CODING SEQUENCE, AND
CC      SOMATIC MUTATIONS IN COLORECTAL TUMORS ARE FURTHER CLUSTERED IN A
CC      PARTICULAR REGION, CALLED MCR (MUTATION CLUSTER REGION). (3) MOST
CC      IDENTIFIED POINT MUTATIONS IN THE APC GENE ARE TRANSITIONS FROM
CC      CYTOSINE TO OTHER NUCLEOTIDES. (4) THE LOCATION OF GERM-LINE
CC      MUTATIONS TENDS TO CORRELATE WITH THE NUMBER OF COLORECTAL POLYPS
CC      IN FAP PATIENTS. INACTIVATION OF BOTH ALLELES OF THE APC GENE
CC      SEEMS TO BE REQUIRED AS AN EARLY EVENT TO DEVELOP MOST ADENOMAS
CC      AND CARCINOMAS IN THE COLON AND RECTUM AS WELL AS SOME OF THOSE IN
CC      THE STOMACH.
CC      -1- SIMILARITY: CONTAINS 7 ARM REPEATS.
CC      -1- DATABASE: NAME=APC; NOTE=Information about APC mutations;
CC      WWW="http://perso.curtie.fr/thierry_sousisl/APC.html".
CC      -----
CC      Query Match      3.1%; Score 258.5; DB 1; Length 2843;
CC      Best Local Similarity 17.9%; Pred. No. 0.00011;
CC      Matches 304; Conservative 238; Mismatches 618; Indels 535; Gaps 69.
CC
QY      30 SHLPEEEKIILAVMDRCKKEBEKQSLKIKEKKAQPTQWFPFSGITELVNNLQPOQ 89
DB      1012 NHMDNDSELTPTIYSLKYSDEQ---LNSGRSPSONENERARKPHILE--DEIKOSQO 1065
QY      90 KQPKNEKEPQTLHOEFEMKQYKMGGESEO---QOQEGADPAPCGIICHTKFAFGCGH 146
DB      1066 RQSNRQS-----TTPYPTTESTDDKHLKRPHPHQOE-----CVSPRYRGAANGST 1112
QY      147 NCSTYQCTKFCARCGGRVSLRSNKKVMVNCNLCKQOELLTKSGAMFYNSGNTLQOPDKV 206
DB      1113 N-----RVGSNHGINVQSGLC-----QEDYEDDKPTNYSEKYSSEQHEE-- 1155
QY      207 PGLRNEAPQEKKALKEQPOFGAPDLSVPAVEKGAHGLTQDPIKNSGVKHOIA 266
DB      1156 -----ERPTNYSLIKIEKRRHNDQPIDYSL-----KTA 1184
QY      267 SDMPDRKRSPSVSHQDQRRRYEEOSEEREDYSQYVSDGMPSPSDVADRRSQREPOFE 326
DB      1185 TDIPESQKQSPFSKSSSGQSSKTE-----HMSSESNITSPBSNMKRONQLPSSAQ 1237
QY      327 EPGHLNYSRNSRGRHSKE---YVDD-----EDVE 355
DB      1238 SRSGQPKAKATCKVSSINQETIQTYCVEDTPICFGRSCSLSSLSAEDEIGCNGTQOEA 1297
QY      356 SRDEYERQREBYOARKRDPNLARYVYKPYTYEOMRIHAEVSRARERHSDVSLAN 415
DB      1298 SANTIQIAIEIKIGTRSAEDPVSEVPAVSOHPRKSSKLGQS--SLSSRAHKKAVERFS 1356
QY      416 ALEDESRISLMDRP-----SRQSVERRAMENORSYMERPREAO--- 459
DB      1357 GAKSPKSGAQTPKRPPEHYQVETPLMSKCTSVS---SLDSFESRSTLAHSVQSPCGC 1412
QY      460 -----GQSSYVQRTSNHSPTPRRSPFLDPRDMKRADSLRKHHLDPSSAVRTKREK 513
DB      1413 MVSGLITSPSDLPDPSGQMPBPERSKTPPP-----PQTA--QTKR-- 1451
QY      514 METMLRNDLSLSDQESVRRPPPRHKSCKGGMQVSLSSSEEL-----ASTPEYITSC 568
DB      1452 ---VPKNAKPALEKRES-----GRKQAAVNAVORVQLPADRTLHRAETSTPDGFCG 1502
QY      569 D-----DEVL-----ESESVEKSGSQ--KGRKRT---SEQG 595
DB      1503 SSSLSALSLDEPFIOKDVLRIMPVYQENDNGNTEPESDPKFNSNENQKEAKETITDSE 1562
QY      596 VLSDSN-----TRSEROKRMYYGHSLSLEEDLEWSEPOIKDSGVDTCS 639
DB      1563 ILDDSDDDDIETILECIIISAMPTKSSRAKKAQPAQASKLPVPVARKPSQL-----PVYK 1616
QY      640 TTLNEHSHSKNHPVTWQPSKDGDL---IGRIILN-----KRLKDSVPRDSCAMGL 690

```

| | | | |
|-----------|---|---|---------------|
| Dh | 1617 | LLSOGNRLQOPCKH-VSPFPGDDMPHYVCQEPINRSTFSTLSJDLIESPPNPLAGEV | 1673 |
| QY | 691 | KVVGSKMTESGRLCAPITKVKKGSGLADYVGHLPBGDEVLEWNGRLLOQATFEVYNTILE | 750 |
| Dh | 1676 | R--GG--AOSG-----EFEKRDITPTEG--RSTDEA-----OGKTSVITPILD | 1714 |
| QY | 751 | SKPEQVELYV-----SRIGD-----LPRIPDSHAOLESSSSPEOSKMRPSISVT | 799 |
| Dh | 1715 | DNKAEGDGLAECINSAMKRGSKHKPFYKKTIDQYO-QASASSAPANKNQLGKKKPT | 1773 |
| QY | 800 | SPMSGELRDYPOFLSGQSLIKLWFDKVGHOLVITLGAADLPREDGRRPNRYVKIYFL | 859 |
| Dh | 1774 | SPVKP-----IPQNTNEYRRV-----RKNADSKNNLNERVF | 1805 |
| QY | 860 | PDRSDKNRRKTYVKKLEPK-----WNQTIYSFVH-----RR | 893 |
| Dh | 1806 | SDNKSCKONLKNNSKDFNDKLPNNEDRVGSFAPSPHHYPIEGPYCFSRNDSLSTL | 1865 |
| QY | 894 | EFRKMLETTLMDQAVRRESEFELGELLILETALLDDEPHYKTIQTHVSSPLPR-- | 951 |
| Dh | 1866 | DFDDDDVDLSR-EKALRKAKENKSEAKVTSHTLELTSNOASANKTQA--IAKOPINRGQ | 1922 |
| QY | 952 | PSPYLPRROL-----HESPRFROR-----SKRISDEYDVDC----- | 987 |
| Dh | 1923 | KPPILOKOSTFPOSSKDIPDRGAIDDEKLONFATIENTPVCFSFNSSLSLSDIDQENNK | 1982 |
| QY | 988 | -----DGVGVSDYRHNG-----RDLQSSLTLSYPEQVMSSNHC | 1020 |
| Dh | 1983 | ENEPKETEPPDSQGPSPKPOASGVAPKSPHVEDTVPVCFSRNSSLSTLSIDSEDLLOEC | 2042 |
| QY | 1021 | SPSGSPHRYD---VIGRTSMSPSPAPPOANVEQGRGTRATHYNTISM-D-RHRMDD | 1076 |
| Dh | 2043 | ISSAMPKKKKPSRLKGDNKHSB-----RNM-GGILGEDLTLDLIDQRPDSEHGSPD | 2095 |
| QY | 1077 | HYSSDRDRDCEAD-----RQ-----PYHRSKTE | 1101 |
| Dh | 2096 | SENFPMKAIQOEANSTVSSLHQAAAACLSRQASSDSDSLTLKSGISLGSFHLTPDQE | 2155 |
| QY | 1102 | QRPLL-----ERTTTRSRSSEBPTN----- | 1122 |
| Dh | 2156 | EKPFTSNKGPRIILKPOEKSTLETKKLESBKGIKGKKVYKSLITGKVRNSEISGCMQ | 2215 |
| QY | 1123 | -LMRSMPSLMTGRSAPSPALSRSHRTGSV-----QTSBSTPFGTGRGROLPLPR | 1174 |
| Dh | 2216 | PLQAMMPSISRRTMHIHGVRNSSSTSPVSKKGPPLKPAKSPSEG---QTATTSP | 2271 |
| QY | 1175 | KGLERSANDIEDRNOMKLNKYQVAGSPRLDEQYHSKYRSG--WDHHRGAD----- | 1226 |
| Dh | 2272 | ROAKRSVWSELSPVAHOT-----SQGGSS-----KASRSGSRDSTSPRPAQPLSR | 2319 |
| QY | 1227 | TVSTKSDSDVDVAVSARTSSASRSTSYMSVQSEBRPGNKRISVTEKMONRQMGVS | 1286 |
| Dh | 2320 | PIQSPGRNRSIPGRNISPPNKLSQLPRITSSPTASTKSSGSGMS-YTS--PGRQM--S | 2374 |
| QY | 1287 | GKMLTKSTISISGDMCSLEKNDGSOPTAVGALUTSGCKKRRSSIGAKWVAIVGLSRKRSRA | 1346 |
| Dh | 2375 | QONLTKQTLKSNASSIIPRSESASKGLNOMNONGANKK-----VELSRMSSTK | 2423 |
| QY | 1347 | SQLSQTEGGGKLL-----RSTYQKSTETGLAVEMKRMNTQOASREST----- | 1388 |
| Dh | 2424 | SSGSSSDRSERPELVLRQSTFIKEAPSPTLRLKLEASAFESLSPSSRPASPPTRSQAOTPV | 2483 |
| QY | 1389 | -----DGSMSYSS | 1397 |
| Dh | 2484 | LSPSLPDKMLSTHSS | 2498 |
| RESULT | 6 | | |
| TPR_HUMAN | | | |
| ID | TPR_HUMAN | STANDARD: | PRT: 2349 AA. |
| AC | P12270: | | |
| DT | 01-OCT-1989 (rel. 12, Created) | | |
| DT | 01-OCT-1996 (rel. 34, Last sequence update) | | |

01-OCF-2000 (Rel. 40, Last annotation update)
 DE NUCLEOPROTEIN TPR.
 GN TPR.
 OS Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 NCBI_TaxID:9606;
 RN [1]
 RN SEQUENCE FROM N.A. PubMed-1437155;
 RX MEDLINE=93064711;
 RA Mitchell P.J., Cooper C.S.;
 RT "The human tpr gene encodes a protein of 2094 amino acids that has
 extensive coiled-coil regions and an acidic C-terminal domain.";
 RT Oncogene 7:2329-2333(1992).
 RN [2]
 RN REVISIONS AND CHARACTERIZATION
 RX MEDLINE=95096166; PubMed=7798308;
 RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
 RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
 RT "tpr, a large coiled coil protein whose amino terminus is involved in
 activation of oncogenic kinases, is localized to the cytoplasmic
 surface of the nuclear pore complex.";
 RT J. Cell Biol. 127:1515-1526(1994).
 RN [3]
 RN SEQUENCE OF 1-142 FROM N.A.
 RX MEDLINE=88262257; PubMed=3387099;
 RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;
 RT "tpr homologues activate met and raf.";
 RT Oncogene 2:617-619(1988).
 CC -1- FUNCTION: COMPONENT OF THE CYTOPLASMIC FILIBRILS OF THE NUCLEAR PORE
 COMPLEX IMPLICATED IN NUCLEAR PROTEIN IMPORT. ITS AMINO TERMINUS
 IS INVOLVED IN ACTIVATION OF ONCOGENIC KINASES.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE
 COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH
 TPR-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER
 COMPONENTS, INCLUDING P62.
 CC -1- TISSUE SPECIFICITY: HIGHEST IN TESTIS, LUNG, THYMUS, SPLEEN AND
 BRAIN, LOWER LEVELS IN HEART, LIVER, AND KIDNEY.
 CC -1- DISEASE: INVOLVED IN TUMORIGENIC REARRANGEMENTS WITH THE MET, TRK
 OR RAF GENES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X66397; CAA47021.1;
 DR EMBL: Y00672; CAA66881.1;
 DR PIR: S00928; S00928.
 DR MIM: 189940;
 KW Heptad repeat pattern; Coiled coil; Proto-oncogene;
 KW Chromosomal translocation; Nuclear protein; Transport.
 FT DOMAIN 78 360 COILED COIL (POTENTIAL).
 FT DOMAIN 422 571 COILED COIL (POTENTIAL).
 FT DOMAIN 575 628 COILED COIL (POTENTIAL).
 FT DOMAIN 758 805 COILED COIL (POTENTIAL).
 FT DOMAIN 834 869 COILED COIL (POTENTIAL).
 FT DOMAIN 934 979 COILED COIL (POTENTIAL).
 FT DOMAIN 1004 1064 COILED COIL (POTENTIAL).
 FT DOMAIN 1138 1166 COILED COIL (POTENTIAL).
 FT DOMAIN 1196 1241 COILED COIL (POTENTIAL).
 FT DOMAIN 1262 1304 COILED COIL (POTENTIAL).
 FT DOMAIN 1354 1434 COILED COIL (POTENTIAL).
 FT DOMAIN 1476 1595 COILED COIL (POTENTIAL).
 FT DOMAIN 527 530 POLY-SER.
 FT DOMAIN 1833 1836 POLY-GLU.
 FT DOMAIN 1957 1964 POLY-ASP.
 FT DOMAIN 2295 2298 POLY-SER.
 SQ SEQUENCE 2349 AA; 265600 MW; AFDD6895CEDCAC9EF CRC64;

Query Match 3.18; Score 255.5; DB 1; Length 2349;
 Best Local Similarity 18.14; Pred. No. 0.00012;
 Matches 341; Conservative 268; Mismatches 675; Indels 599; Gaps 76;
 10 RPAPTASOPPPPEMDLSHL-TEERKIIILVMDROKKEEKGSLKIFENKAP 68
 644 RPSTQIVSTPAPVPIESTAIKAKALQOLDFEYKPKKKAENE----- 690
 69 TQWPFSGITLVNVLQPOQKQNEKEPQ-TKLHQ-----PEMKKEQKKMGE- 118
 691 -----KIQNEOLEKLEQVYVLDLSQNTKISTQDFASKREYEMQDNVEGRREIT 740
 119 SOOOOEKQADPATICIGIKTKFADCGCHNCSCYQKFCARCGRVLSNKMVMVYCNLCR 178
 741 SLHERNQLKLTAT-----Q 754
 179 KOQELITKSGAMFYNSGNTLQDPQKVP-RGLRNEEAQOEK-----KAKLHEQPOFG 231
 755 KQEQII-----NTMQDLRGANEKLAVALVRAENLKKKEEMKLSEVRLSQORE--- 803
 232 APGDISVAVEKGRAGHGLTRDTI-----KNGSGVKHQAISMPDRKSPSVSDQNR 286
 804 -----SLAEORGQNLILTNLQTLQIGLERSETETKQRLSSQI---EKLHEISHLKKKL 855
 287 YEOSERE-----DYSQVPSDGTMPSPSPDYADRSQREPOFEYEEGHILNYRD 335
 856 ENEVQDRHITLNLNLDVQLDTKQDLTETNLHLNTEKLKNAQKEIATLKHLSNMEVOY 915
 336 SNRGRHRSKEYIVDEDEYSEDRERQREBEYQARY---SDPLARYPV----- 384
 916 ASQSQRTGCKGQPSNKEDVDLVSQRLQTEQVNDLKERIKTSTNVEQIOAMVTSLEFS 975
 385 --KQPYEQRMTIAEVS-----RARHRRHSDVSLANAELEDSR-----ISL 425
 976 LNKQVYTEVKNIEVRLKESAEPTOLEKLMVEKKEQELQDDKRAIESMEQOLE 1035
 426 LMDPRPSRQSVSERRAAMENQSYSEMTREAOGOSTPQRTSNHSPPRPSPIPLDR 485
 1036 LKLTSSVQNEVOEALQRASTALSNQQAARDQCOEQAIAVEAQN---KYERELMLHA 1090
 486 PMRRADSLRQ-----HHDPSSAVRKT-----REKKM-----E 515
 1091 ADVEALQAAKEQYKSNKASVROHLETTQAKESQULECKASWBERERMLKDEVSCVGRCE 1150
 516 TMLRNDLSQDSQSVRPPPRPHKSKGKMROVSLSSSE----- 557
 1151 DLEKQRLHLDQLEKL---SDKYVASVKEGVQGLNLSLEEGKSQQLLEILLRIRREK 1207
 558 -----ELASTPEYISCDVLESESVSEKQDSQKKR-----KTS 592
 1208 EIAETFEVAVQESLRYRQREVELLERLOLEDSLNAEREKVQYATKMAOHELMKTE 1267
 593 EGVVLSQNSRSEKQKRMVYGGSHLEDELEWSEPOTKDSGVDTCSSTTLNEHSHS--- 649
 1268 TMYVMEETNMLKEERK-----LEQDILQOMQAKYRKLELDILPLQEANAELESEKSGM 1320
 650 -----DKHPVTQPSKQDRLIGRLILNKLKLDGVSVPDSDGAMGLK 691
 1321 LQAEKKLLEEDVKKRKARNHIVSQ--KDPDTEYRKLSEK-----EYHTK 1366
 692 VVGKMTESGRICAFILTKVKKG-----SLADYVGHLPBGDEVLEWLN----- 732
 1367 RIQQLTEIGRLKAEIKASNSALNNONLLOSLEKDLKVKETIKEDKIDAKIIDQ 1426
 733 -----GRLLQ-----GATFEVYVNIILSKPEPELVLSRPIDIPRIPIST 775
 1427 KVKITTVQKKIKGRYKQYELKMAQDQVMETSAOSSGDHDEQHVQEMOEL---KET 1482
 776 HAQLESSSSSFESQ-----KMDRSISVTPSMSPGMKRDVQFQLSGO-- 817
 1483 LQAETKSKLSQSVENLQKTLSEKETEARNLQEQTVQLOSEL---RLRQDLQDRITQPE 1540


```

Db 1619 OSRLQAOAHVSTPGDDVP-----RYVCYEGTPIINFSTATSLSDLTIESPPNLEAGDC 1672
OY 161 GYLSLSNKNVWVWVNCJCRKQOEILTKSGAFYNSGNTLQDPQKVPBGLINNEAPQEK 220
Db 1673 VRASVSGE-----EKMDIIPTEG-----RSTDEAQGK 1702
OY 221 AKLHEOPQOGA---PQDLSV---PAVEGHAHLTR-----QDTIKNGSV-KHQ 264
Db 1703 VSSIALPDLIDGSAEEDDILAECLNSALPKGRSHKPFVKKIMDOVQASMTSGTNNK 1762
OY 265 IASDMSDBKSPSVSDONRRYE-OSEEREDYQYVSDGTMPPSPEDYADRSGQREPQ 323
Db 1763 I--DTKKKKTSPYKPPQPTETRTKAKNTDSKVNNTEET-----FEDNDSKKQS 1813
OY 324 FYEPGHLNRYDSNRGRHRSKEYIVDEDEVSREYERQREBEYQARYSDPMLARYP 383
Db 1814 LKNNPDNLN-----DKLPDNE-----RVRGGFEDSPHHYAP 1846
OY 384 VKPOPIEOMRIHAESRAHERHSDVSLANAFLEDSRISLRLND---RPSRQSVGR 440
Db 1847 IEGTPY-----CFSRND-----SLSLDFDDDDVDLSREKAEIRKGEKSDSA 1890
OY 441 RAMENORSYMERTRRQAGSSYPQRTSNHSPPTPRSPPLP-----DRPDMRADSL 494
Db 1891 KYCHTPSSSQQARKKAQASTKHP---VNRGPKPPLIQEQPTFPQSKDVPDRGAATDE 1947
OY 495 RKOHLDPSSAVRKTREKMETMLRNDLSL-----DOSESVRPPPPRHKSKK 543
Db 1948 KLOFAIENFTPV-----CFSRNSLSLSLDVDOENNNNEETGPYADAP-ANAQO 1997
OY 544 GGMROVSLSSSEELASTPEYTSQDDVLESESYSEKDGSKGKRTSEQVLSDSNTR 603
Db 1998 GKPKQASGVAKSFHVEDTPVCFSRNS-SLSLSLTD-----SEDLLEKCISS 2044
OY 604 SEROKKMYGSHLEDELEWSEPO-----IKDSGVDTCSTLTNEHSHS-DKHPTVM 656
Db 2045 AMPKRR---PSLKLKGGEWQSPKVGSVLAEDLTLLDKIQRESEHGLSPDSENFDM 2100
OY 657 OPSKQDRLICRILINKRLKKGSPVR---DSGAMLGKLVGKGKTESRCAFITTKYVK 712
Db 2101 KALOEGANSIYSLHOAAAAACLSRQASSDSDSLSLK-----SG-----VSL 2144
OY 713 GSLADYGHLEHGEDEVLEWNG---RLQGAATFEFVYNLLESKEPEQVELVSRIGDI 768
Db 2145 GS---PFHLTPDDEKFTSHKGRILIKPE---KSTLEAK----- 2179
OY 769 PRIPDSTHAQLESSSSSFESQK-----MDRPSIVTSPMSPQMLRDVPOQLSGQLS 819
Db 2180 -----KIESENKGIKGGKVKYKSLITGKIRSNSEISSOMKOPLOTNMPSISGRFTM 2230
OY 820 IKLMPDKYGHOLIYIILGAKDLPSEDEGR-RNPYVKIIFLDRSDXNKRRRK-TYKTL 877
Db 2231 I-----HLPGVANSSSTSPVSKGPPKLPKTPASKSPSECPVATTSRGTKPAVKSEL 2282
OY 878 EPKNMOTFIYSPVHRRERERMLFITLMDQAVREBESEFLGELIETALTALIDEPHMY 937
Db 2283 SPITRQTHIGSKNKGPRSGSRDST---PSRPTQ----- 2315
OY 938 KLQTHDVSSLPLPRSPYLPRLQOL---HGESPTRRLOQSKRISDSEVDYDCEDGVY 993
Db 2316 -----PLSRPMQSGRNSISPRNGISTPNKLSQLPRTSSPSTASTK-SGSGMK 2364
OY 994 SDYRNGRDLOSSTLSVBEQVWVSNHCSPSGSPHARDVIGRTFSWSPSAPPPQRYNEQGH 1053
Db 2365 S-YTSPGQLOQOONLS-KQTGLSKN---ASSIP-----RSEASAGLNMNNSNG- 2409
OY 1054 RGRATGYHNTSRMDRHRVMDHSSDRDRDCEADROPYHRSRSTEQPFLLERTTYS 1113
Db 2410 -----SNKKVELSRM-----SSTKSSGSE-SDR-----SERPALVQGSTFL 2444
OY 1114 RSSERPDTNLMSKPSLMTGSAAPS-----PALSRSHPRGSGVOTSTPSSTP 1160
Db 2445 K--EAPSPILRRKLEESASFESLSPSSRPDSPTRSQATPVLSPLPDM-SLSTHPSVOA 2501

```

```

OY 1161 CTRGRGROLPOLP--KGLTFRS-----AMDIERNQMKLKYKQVAGSDPRL 1208
Db 2502 GGMK-----LPPNLSPITEYSDGRPSKRHDIAHSHPSPRLPVNR---AGTKR-- 2549
OY 1209 ODHSKYSRGMDPHRGADVTSTKSSDSDVDVSAVSRSTSASRFSSTYSVSGSERPRGN 1268
Db 2550 ---EHSKH-----SSSLPRVSTWTRTSGSSSLTASPSSESEKAKSEDE 2589
OY 1269 RRTSYFTSMQNRQMGVSGKNLTKSTISGDMCSLEKNQDSQDITVAGALGTSGKRRSS 1328
Db 2590 KHVNSVPCPRQMKENQVPR-----GTWRKIKESETISPTNT-VSOTTSQ---AA 2635
OY 1329 IGAKVAVIIGLSKRSASASOLQTEGGCKLIRSTVORSTETGLAVE---MRNMTRQASR 1385
Db 2636 SGASEKTL-----YQMAVAVRTEBWWIRIDCPLNNRSGSRPT 2676
OY 1386 ESTDGSMSNSYSEGNLIFPGVRLASDQSPDFLDGLGPAOLVGRQTLATPAMGDIOVGM 1445
Db 2677 GMPPIYDISISEKN---PSIKSKDTPQCKOSVSGSPVQVGLENRL----- 2721
OY 1446 DKKGLEVEILIRARGLVVRGSKTLRPAYKYVLLDNGVCIARKTKTVARKTLEPLYQL 1505
Db 2722 --NSFIQVEAPDEQKGTETAGQGS-PAPVAET---GETMAER-----TPRESS 2765
OY 1506 LSPFESPOGRVLIQTIWGDVGRMDHKSPMGVAOILDELFLSNVATGWFKLPSPSLVDP 1565
Db 2766 SSKHSSPSGTVARVYPPVNNPSPRKSADSTSA-----RQSQIFTP 2807
OY 1566 TSAPLTRRASQS-SLESSTGPSYSR 1589
Db 2808 VGSSTKRRDSTSTESSGAOSPKR 2832

RESULT 8
APC_MOUSE STANDARD; PRT; 2845 AA.
ID APC_MOUSE
AC 061315; Q62044;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ADENOMATOUS POLYPOSIS COLI PROTEIN (APC PROTEIN) (MAPC).
GN APC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A. (ISOFORM 1 AND 2), AND VARIANTS.
RC STRAIN=C57BL/6J, AND CAST/EI; TISSUE=Brain;
RC MEDLINE=92263101; Pubmed=1350108;
RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
RA Luongo C., Gould K.A., Dove W.F.;
RT "Multiple intestinal neoplasia caused by a mutation in the murine
RT homolog of the APC gene."
RT Science 256:668-670(1992).
RN [2]
RN ERRATUM.
RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
RA Luongo C., Gould K.A., Dove W.F.;
RL Science 256:1114-1114(1992).
RN [3]
RN SEQUENCE OF 1-45 FROM N.A.
RC STRAIN=BALB/C; TISSUE=Liver;
RA Dicker F., Lambertz S., Reilmair A., Ballhausen W.G.;
RT "The murine APC gene: alternative splicing of 5' untranslated
RT region segments."
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [4]
RN ALTERNATIVE SPLICING.
RX MEDLINE=94061824; Pubmed=8242607;
RA Oshima M., Sugiyama H., Kitagawa K., Takeo M.;
RT "APC gene messenger RNA: novel isoforms that lack exon 7."

```

RL Cancer Res. 53:5589-5591(1993).
CC -1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-
CC CATEININ. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE
CC ALLOWS THE DOWNREGULATION OF CYTOPLASMIC BETA-CATEININ (BY
CC SIMILARITY).
CC -1- SUBUNIT: FORMS HOMODIGOMERS AND ASSOCIATES WITH CATEININS (BY
CC SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS: 1 (SHOWN HERE), 2, 3 AND 4; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN LIVER, SPLEEN, KIDNEY, HEART,
CC LUNG, BRAIN, STOMACH, INTESTINE, TESTIS AND OVARY.
CC -1- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 7 ARM REPEATS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M88127; AAB59632.1; -
CC EMBL: U02937; AAA03443.1; -
CC HSSP: Q02248; 2BCT.
CC MGI: 88039; APC.
CC InterPro: IPR000225; -
CC Pfam: PF00514; Armadillo_seg; 4.
CC DR POSITIVE: P550176; ARM_REPEAT; 1.
CC KW Anti-oncogene; Phosphorylation; Alternative splicing; Repeat;
CC Coiled coil.
CC FT DOMAIN 1 61 COILED COIL (POTENTIAL).
FT DOMAIN 125 245 COILED COIL (POTENTIAL).
FT REPEAT 1 728 LEU-RICH.
FT REPEAT 451 493 ARM 1.
FT REPEAT 503 545 ARM 2.
FT REPEAT 546 589 ARM 3.
FT REPEAT 590 636 ARM 4.
FT REPEAT 637 681 ARM 5.
FT REPEAT 682 723 ARM 6.
FT REPEAT 724 765 ARM 7.
FT DOMAIN 739 2834 SER-RICH.
FT DOMAIN 1130 1156 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 1556 1575 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 1864 1891 HIGHLY CHARGED.
FT VASPLIC 243 276 MISSING (IN ISOFORM 2 AND ISOFORM 4).
FT VASPLIC 310 410 MISSING (IN STRAIN CAST/EI).
FT VARIANT 120 120 V->A (IN STRAIN CAST/EI).
FT VARIANT 493 493 V->I (IN STRAIN CAST/EI).
FT VARIANT 797 797 Y->F (IN STRAIN CAST/EI).
FT VARIANT 1330 1330 A->T (IN STRAIN CAST/EI).
FT VARIANT 1618 1618 A->S (IN STRAIN CAST/EI).
FT VARIANT 2294 2294 G->A (IN STRAIN CAST/EI).
FT VARIANT 2496 2496 H->Q (IN STRAIN CAST/EI).
FT VARIANT 2523 2523 T->A (IN STRAIN CAST/EI).
FT VARIANT 2813 2813 T->S (IN STRAIN CAST/EI).
SQ SEQUENCE 2845 AA; 311086 MW; 145CA73CF570AA99 CRC64;

Query Match 3.1%; Score 254.5; DB 1; Length 2845;
Best Local Similarity 18.8%; Pred.No. 0.00017;
Matches 362; Conservative 241; Mismatches 692; Indels 633; Gaps 86;

QY 45 DROKKEKEKESV---LTIKEF--HKAQPTQW-----PPSGITTELVNANV 84
DB 1147 EEEHEHEEHEEPTNYSIKYNEKHHVDPIDSLKYADISSQKPSFSKNSAOST- 1205
QY 85 LOPPOKOP---NEKEPOTKLHOQFEMKYKKMGEESSQOQEOEKDAPTC--GICHT 138
DB 1206 -KPEHLASSSENTAVPPNNAKRONLRSSAQRNG-----QTKGK--TTCKVPSINQET 1256
QY 139 KPADCGHNCASYC--QTRFC-ARCGGVSLSRSNKVMVNCILRKQOELLTKSGAWFVNSG 195

DB 1257 -----IOTVCVEDPTPICFSSCSSLSLSSADDEIGCDQTTQTEAD-----S 1296
QY 196 SNTLOOPDOGVPPGLRNEEAPOEKKAKLHQPGQGAAGDLVPAVEKG-----RAH 247
DB 1297 ANTLQTAEVKENDVTKSAEDPATE-----VAVSONARAKSRLOAS 1338
QY 248 GLTRQDTIKNGSGVKHIOIASDMPDRKRSVSRDQNRKRYDQSEEREDYSQVPSDGTMP 307
DB 1339 GLSESEY-----RHKKAVFSSGAK-SPSKSGAQT-----P 1368
QY 308 RSPSDIADRRSQREPOQVEEPGHILNTDSNRGRHRSKEYIVDEDEVESRDEYERQREE 367
DB 1369 KSP-----PEHYVOETPLVFSRCT-----SVSLDSFESRSIAS 1402
QY 368 EXQARYRS-----DENLARYPK---PQYEQMRILHAEVSRAR---HEER 407
DB 1403 SVQSEPCSGAVSGIISPSDLPDSFGQTMPSPRSKTPPPPTQVAKRKVPKPAAKR 1462
QY 408 HS--DVSLAAAELE-----DSRISILRMDRSRORSVER 440
DB 1463 ESGPKQTVANNAVQVRQVLDVDTLLHFATFESTPDGFCSSLSLALSDDEFIQLKVELR 1522
QY 441 RAAMENQRYSMERTBAQCOSSYDQRTSNHSPTPRRSPIPLDRPDN----- 488
DB 1523 IMPPVQENDNGNETESQPESENQKEVEKPDSEKDLDDSDDDIEILECITISAMP 1582
QY 489 -----RRADSLRKQHHLDPPSAVVRKTRREKMETML-RNDSSLSDQSESYRP--PPPR 537
DB 1583 TKSSRKAKKLAQYASKLPPRYAKRPQSLPYKLLPDAQRLOAQHVSFTPDQVPRVYCV 1642
QY 538 ---PHKSKGKGKMPROVSLSSSEELAS-----TPREYSCDVELESESVSEKGSQ 585
DB 1643 EGTPINSTATSLSDLTIESPNNELATGCVGRAGIQSEFEKRPDIPIEGSTD--DAQ 1699
QY 586 KGR-----RKTSQGVLSU-----SNTRS--- 604
DB 1700 RGRISIVTDPDLDNKAEEGDIILAEICINSAMPKKGSHKFPVKKIMQVQOASSTSGAN 1759
QY 605 -----ERQKKRMYVGHSLLEDLEWSEPOIKSGVDTCSTILNEHSHSD---KHPTW 656
DB 1760 KNOVDYTKKKKPTSPVKPMPQTEYRTYVRK---NTDSKVNWTEETPSDNKSKKPSLQ 1815
QY 657 QPSK-----DGDRLIGRILLNK----- 673
DB 1816 TNKAKFEKLPNNDRVRGTFALDSPHHIYPIEGTPYCFSNDSLSLDPDDDDVDSRE 1875
QY 674 --RLKDGSVPPDSGAMLGK-VVSGKMTESGRLCFTYKKKSLADTVGHLRPGDEVLE 730
DB 1876 KAEILRKESKESDEBAKYTCRPEPNSOQAAKSOASIKHPANRAQSKPVLOKQFPQSS 1935
QY 731 WNGRLLAGATFEFVYNNIILESKRPEQVELVYSR-----PIDDIPRIIPSTNAQLESSSS 785
DB 1936 KDGDRCGAATDEKQNLATIENTP-----VCFSRNSSLSDSIDID-----QENNNKE 1982
QY 786 FESQKMDRPSISYVSPMSPGM-----LRDVPOFLSGOLSIKLWF---DKYGHOLI 832
DB 1983 SEPIKAEAPNANSQEPERKQASGYAPKSFHVEDIPVCSRNSSLSSLSIDSEDDLQGC 2042
QY 833 VTIIGAKDLPSREDGRPPNPVYKTYFLPDRSDKNKR-----TKYVKLTLEPKWN 882
DB 2043 SSAMPKRRKRSR-----LKSESEKQSPRKVGGIILAEIDLTLDKLDQRPSE 2088
QY 883 QTF-----IYSPVHRERERENLETLTDQARVREDESEFLGELLIE 924
DB 2089 HAFSPGSENEWDKAIQCGANSTVSLHQAAAAACLS-----ROASSD--SDSLIS 2137
QY 925 LETALLDDEPHWYKLQTHDVSSLPL-----PRSPYLPRLRQIGESTPRLQSRK 974
DB 2138 LKSGISLIGSP--FHL-TPDQEKEPFTSNKGRILKPGKSKLLEAKXI--ESNNKIKKGK 2192
QY 975 RI-----SDSEVSDYDCDEGVGVSYDRIHNGRDLQSSSTLSVPPQVMSNHCSP-- 1022
DB 2193 KYVKSILITGKIRSNSEIS--OMKOPLPITNMPISIRGRFTM---IHIPGLRNSSSSTSPVS 2247


```

Db 925 KFEDEGAFEESEETGDYEEKAEPEEPEDEDEHVCASAKSHSPTE--DEESAKAEAD 983
QY 324 FYEPBGHNTYDSNRGRHGRHKEYIVDEDEDVBSRDEYEQ-----RREDEYQARYSDSN 378
Db 984 AYIREKRESVASGDRAEDMEDEALEKEGADEQSEEDAEDEDAEDAREEYE----- 1035
QY 379 LARYPVKPOPEEQMIRIAEYRSARH-----ERRHSDVSLANAELEDSRISLMDRPSHQ 434
Db 1036 -----PEKHADEYVAVYDKAAEAGAEBOYGLTTPYQ-----LGAQSPGAE 1080
QY 435 RSVSERRAAM-----ENQSYSMEERTREAQO-----SSYQRTSNHSPPTPRKSPILDR 485
Db 1081 PASSIHDETLPGCSESEFATASDEENRDOPEEFTATSGYQSTIEIS-----SEPTPMDE 1135
QY 486 PDMRRADSLRQHHLDPSAAVARKTKREKMETLRLNDSSLSDQSESVRPPRPHKSKGG 545
Db 1136 -----MSTPRDVMASDETNNETESPQOE----- 1159
QY 546 KMRQVSLSSSEELASTPEYTSODDY-ELESESVSEKSGDSKGRKTSBOGLSDNTRS 604
Db 1160 -----VNTIKYESSLYS-QEYKRPADYPTLNGSESGKTDATGCKOYNASASTISPS----- 1211
QY 605 ERQKKRMYTGGHLSLEEDLEWSEPOIKDSGVDTCS-----STLINEHSHSKDHPVTWQPSK 660
Db 1212 -----SMEDD-KFSKRALRDA---YCSEVKAATTLIDIKDISAVSEKVSFG 1255
QY 661 DGDRLIGRLILKRLKDSVPRDSCAMGLKRYVGKMTESGALCAFITVKKGSLADYTG 720
Db 1256 -----SPSLSPSPSPPLEKTPLEGRSVNFLTPN----- 1284
QY 721 HLRPGDEVLEWNGRLQGATFEENVYNIILESKPEPEVLEVRPIGDIPRIDPSTHAOLE 780
Db 1285 -----ELKVAAEAEVAVSEVYQOEYVE 1307
QY 781 SSSSSFEQSKMD--RPSISVTSMPSPGMLRDVPOFLSGOLIKLMDKYGHOLIYTLIGA 838
Db 1308 EHCAASPEDKTLEVVSPSQSVT-----GSAGHTPYQSPPT-----DEKSSH----- 1347
QY 839 KDLSREGRGRNRPVYKTYF-LPDSDKNKRKT-----KTVKKTLEPKNQTF 885
Db 1348 --LPLEVLEKP--PAVPYSPFESDADKDENERASVSPMDEPVPDSPIKIVLSPFLKSPPL 1403
QY 886 IYSPVHREPREBMLEITLMDQARVREESEF-----LGEILIELETLALDD- 932
Db 1404 IGSSEAVSYF-----LSADDKASGRGASPEEKSCKOGSDQVSPVEMTSTLYQDK 1457
QY 933 -----EPHMYKQTHDVSSLPRLPRSPYLRROLHGESPTR-----RLQ 971
Db 1458 QEGKSTDFAPIKEDRGOKKTDDVEAMS--SOPALALDERKLDDVPTQIDVSGFSFKED 1516
QY 972 RSKRISDEVDYDCED-GVGVVSD-YRANGNDLOSSTLY-----PEQVSSNHCSPS- 1023
Db 1517 TKMSISECTVSDKSAVPEDEGAETJYSHMEGVASVSTASVATSEPEP--TTDDVSPSL 1574
QY 1024 -----GSPHVRVYIGTRMSVSPAPPPQRNVGCGHGRATGCHYNTISRDRHVMDDHVS 1079
Db 1575 HAEVSPSTIEV-----DDSLIS 1591
QY 1080 SDRDRCEADROPYHRSRSTEQRLLETTTTRSSRSEPRDNLNMSMP--SLMTGRSAP 1137
Db 1592 V-----SVUQTPTTFOETEMSPSKKECPRPMSTISPPDFSPKTAKSRT 1633
QY 1138 PSPALSRHPRTGSV---QTSSTSTPGTGRGRQLPQLPRKT-----LERSA 1182
Db 1634 PVQD-HREDEGSSMSTIEGOESPEQSLAND-FSRQSPDHPTAGVLAHTENGTEVDYDPS 1691
QY 1183 MDIEERNOMKLINKYQVAGSDPRLQDYHNSKYRSGWMDPHRGADVSTKSDSDVDVDSA 1242
Db 1692 SDMODSSLSHKT-----PRMEEPSTYQ-----DNDLSELLS 1722
QY 1243 VSRITSSASRFSSTYSVYOSERRPGRN-----RKISVFTSKMOROMGVSGKNLTKS 1293
Db 1723 VSOV-EASPSITSSAHTPSQIASPLQEDTLSDVAPPRDMSLVASTLSEKQVSLGEGKLSR 1781

```

```

QY 1294 TSIIGDMCSELEKNDGSDPAVAGALGTSGKKRS-----SIGAKKVAIVGLSRKSSASOL 1349
Db 1782 SDIS-----PLTPRESSPLYPSTSDSTSAVKEKATATCHSSSSPPIDAASAEPTGERASVL 1837
QY 1350 SQT-----EGCGK---KLRSYVORSTETGLAVEMRMNMTROASREST 1388
Db 1838 FDMQHHLLALNROLSTPGLKEDSGCKTPGDFSYAVQKEEF-----TRSPDEEDY 1887
QY 1389 DGSNNST-----SSEGNLIFPGV-----RLASDSQSFQDLGLGPAQLVGRQTLATPAMG 1438
Db 1888 D--YESYEKTTTSDYGVGYEYKERTTTPKSPSDGYS-----YETIGK-TTKTPEDG 1936
QY 1439 DIQVGMDDKKQGLEVEILIRARGLVVKKSGKTLPPRYVYVILLDNGVCIAKKTKYARKTL 1498
Db 1937 DYSEIIEKTTTPEE---GGYSYDISEKTTSPPEVSGY-----SYETKERSRLL 1984
QY 1499 EPLYOQLSFEESPOG 1514
Db 1985 DDLSN---GYDDSEDG 1997

RESULT 10
NCRL_HUMAN
ID NCRL_HUMAN STANDARD: PRT; 2440 AA.
AC 075376; 09UPV5; 09UQ18;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NUCLEAR RECEPTOR CO-REPRESSOR 1 (N-COR1) (N-COR).
GN NCOR1 OR KIA1047.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=98393736; PubMed=9724795;
RA Wang J., Hoshino T., Redner R.L., Kajiyama S., Liu J.M.;
RT "ETO, fusion partner in t(8;21) acute myeloid leukemia, represses
RT transcription by interaction with the human N-COR/msin3/HDAC1
RT complex.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:10860-10865(1998).
RN [2]
RP SEQUENCE OF 782-2440 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=9937452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K.-I., Hiroseawa M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:197-205(1999).
RN [3]
RP SEQUENCE OF 974-2440 FROM N.A.
RX MEDLINE=99375328; PubMed=10444336;
RA Nagaya T., Chen K.-S., Fujieda M.,
RA Horwitz K.B., Lupski J.R., Seo H.;
RT "Localization of the human nuclear receptor co-repressor (hn-COR) gene
RT between the CM1A and the SMS critical regions of chromosome
RT 17p11.2.";
RL Genomics 59:339-341(1999).
CC -|- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
CC -|- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE
CC ABSENCE OF LIGAND.
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -|- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2

```

AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION DOMAINS (ID1 AND ID2). ID1 CONTAIN A CONSERVED SEQUENCE REFERRED TO AS THE CORN BOX. THIS MOTIF IS REQUIRED AND SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES FLANKING THE CORN BOX DETERMINE NUCLEAR HORMONE RECEPTOR SPECIFICITY.

-1- SIMILARITY: CONTAINS 1 SANT DOMAIN (COMPOSED OF SANT-A+SANT-B).

-1- SIMILARITY: CONTAINS 2 CORN BOX.

-1- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@sib-sib.ch).

EMBL: AF044209; AAC33550.1; -

EMBL: AB028970; BAA82899.1; -

EMBL: AB019524; BAA75814.1; -

EMBL: 600849; -

InterPro: IPR001005; -

Pfam: PF00249; myb-DNA-binding; 2.

PROSITE: PS50090; MYB.3; 1.

DR Nuclear protein; transcription regulation; DNA-binding; Repressor; Coiled coil.

KW

FT DOMAIN 174 216 COILED COIL (POTENTIAL).

FT DOMAIN 254 312 INTERACTION WITH SIN3A/B.

FT DOMAIN 299 328 COILED COIL (POTENTIAL).

FT DNA_BIND 437 482 SANT-A (POTENTIAL).

FT DNA_BIND 625 670 SANT-B (POTENTIAL).

FT DOMAIN 501 557 COILED COIL (POTENTIAL).

FT DOMAIN 607 617 PRO-RICH.

FT DOMAIN 677 687 INTERACTION WITH ETO.

FT DOMAIN 988 1816 CORN BOX OF ID1.

FT DOMAIN 2055 2059 CORN BOX OF ID2.

FT DOMAIN 2263 2267 POLY-GLN.

FT DOMAIN 58 64 POLY-ALA.

FT DOMAIN 593 603 POLY-PRO.

FT DOMAIN 1032 1035 POLY-ALA.

FT DOMAIN 1707 1712 POLY-SER.

FT DOMAIN 1952 1963 L -> V (IN REF. 2).

FT CONFLICT 1014 1014 PP -> SS (IN REF. 2).

FT CONFLICT 1508 1509 W -> R (IN REF. 2).

FT CONFLICT 1561 1561 W -> H (IN REF. 2).

FT CONFLICT 1567 1567 O -> H (IN REF. 2).

FT CONFLICT 1567 1567 O -> H (IN REF. 2).

FT SEQUENCE 2440 AA; 270263 MW; 60A4D7964D00EDAB CRC64;

Query Match 3.1%; Score 253; DB 1; Length 2440;

Best Local Similarity 18.4%; Pred. No. 0.00016;

Matches 32; Conservative 229; Mismatches 618; Indels 614; Gaps 76;

QY 6 GGRGPRAPTPASOPPPPEMDLHLTEERKILLAVDRKKKEEK-EGSVTKIK--- 61

DB 150 GGRGPRAPTPASOPPPPEMDLHLTEERKILLAVDRKKKEEK-EGSVTKIK--- 61

QY 62 ---EEHKAPOTQWPFEGSGITTELVNVLPOOKOPNEKEPOTKLHOQF-EMYKQVYKME 117

DB 207 QOLEEAAKP-----PEPEKPSPPVEQKHRSIVQIITDENKKA 248

QY 118 ESQOOOEOKGAPTCGICHTKFEADCGCHNS---YCQTKFACRCGRVSLSNKAMWVC 174

DB 249 EAH-----KIFEGIGKVELPLTYNCP-----SDTKVYHE 277

QY 175 NUCRQOEILITSGAMFYNSGNTLOQPDOKVPRGL-----RNEAP----- 218

DB 278 NI--TFTNQMRKLLIFFRRNHARQROCKICQRYDQLEAMEKVKVDRLENNPRKAKE 335

QY 219 KKAKLHEQFOFQAGADSLVPAVEKGRAGLTRODTIKNGSV-----KHQIA----- 266

DB 336 SKTEVEYKQF-----PEIKRQROQERFOVQVGRAGLSATIRSEHISIIDG 386

QY 267 -SDMPDRKRSVSRDQNRREYQSEEREDYSQVPSQGTMRPSDYADR-----SOR 320

DB 387 LSEQENNEKQROLSVLRPPMFDAQOR---VKFTNMGLMEDPKVKYKQDFNMVWDH 443

QY 321 EPQYEEFGHLNYSNRGHRHSKEYIVDDEVEDSRD-----EYEQRREVEYQARY 374

DB 444 EKELFKD-----KFTQHPKNGFLIASYLERKSPDVCVLYYLTKKNNKALVR 492

QY 375 -----SDPNLAAYPVKQPEYQOMRIHAESVSRARRHRSHDVSLANAELEDSRISLL 426

DB 493 RNYGRGRGNOQIAR-POQEEKVEEKEDKAKTEKKEEKDEEB--KDEKEDSKENRK 549

QY 427 RMDPRQRQSVSR-----AAMENQSYSEK--TTEACQSSYPORTSHSPR 474

DB 550 EKDKIDGTAETEEREQOATPRGRKTANSQGRKGRITRTMTDEAASAANAATEPP 609

QY 475 TPRSPILDRPDMRADSLRQ-----HHLDPSAVKTKREKMETLNR----- 520

DB 610 PLPPPEPTSTPEVETSRWTEEMEVAKKGLYEHGNMAALAKMGTGSAQCKNFYNY 669

QY 521 -----DSLSDQSESVRPPRPHKSKKQKMYQVSLSSSEBELASTREYSCDVELE 574

DB 670 KRHHMDNL-----QOHKQTSRKPREERDVSQCEASVASTYSAQEDIEDIAS 717

QY 575 SESVSEKDSQKGRKTSF---QGVISDNTSENKRRKKMYGCHLEEDL----- 622

DB 718 NEENENP--DSEVEAVKPSDSEENATSRGNTPEVALEPTTETAPSTSPSLAVSTKPAE 776

QY 623 -RWSEPOIKDS-GVDFCSSTLN-----EEHSHDKHPYTWQPSKD----- 661

DB 777 DSVETQVNDISAEIARQMDVQOCHSAEESVCPRPATKADSVDEYVRPENHASKV 836

QY 662 -GDRLIGRILKLKLDGVSVPDGSAMLGLKYVGKM-----TESGRCLAFITKVKGSL 715

DB 837 EGDNTERDL--DRASEKVEPRDEDLVVAQIQAQREPOSDNDSSATCS----- 884

QY 716 ADYVGLIRGDEVLQENGRLLQCATFEYEVNIILESKP--EPQVELVSRPIG---DI 768

DB 885 -----ADEVDGDEP-----ERQRMFMDKSPSLNLTGSLIVSPLKPNPLD 927

QY 769 PR-----IPDST-----HAQLESSSSSFESQKMDRST----- 796

DB 928 POLQHNAAVTPRWVSCPPNIPICTPVSGYALYQRIKAMHESALIEORORQOIIDLEC 987

QY 797 -SVTSP---MSQGLRDVQFLSGQLSTKLMFDKVGHOILVITLGAKKDLPSREDGRPN 851

DB 988 RSTSTPCGTSGKSPNRWEVLQ-----PADHOLITNLPEGVRLPTTPTRPP 1034

QY 852 PVKITYFLPDRSDKNKRTYVKTLEPKWNOFTIYSPVHRRERERMLTITLDAQARV 911

DB 1035 PLIP-----SKTT-----VA 1045

QY 912 EEESEFLGILLLEVALLDDEPHWYKLOTHDVSSLPRSPY-----LPRROLHGE 964

DB 1046 SEKPSFT-----MGGISISQGTPEY--LTSHNOASYTQETPRKPSVGSISGLPRQESAK 1098

QY 965 SPT---RLQSRKRISDSEVDYDCEDGVGVAGDYRHNG-----RDLOS 1005

DB 1099 SATLPYTKQEEFSPRSONSOE-----GLVRAQHEGVVGRTAGALQSGSTIRGPT 1150

QY 1006 STLSPRVQVSNHNCSPSGSPHRVDVIGTRTSWSPSADPPORNVQOGRGTRAGHYVTI 1065

DB 1151 SKISV--ESIPSLRSGITOGTPA-----LPQGTIPTEALVKG-----SI 1187

QY 1066 SRMDRHRVMDHYSDDROCEADR-----QPYHRSSTEDORPLERTT 1110

DB 1188 SRMP-----TQDS--SPKGRF--EAASKGVHIEGSGHILSTDNINKNAEGTRSPATAEI 1241

QY 1111 TRSSSERPDTNLMKRSNPSLMTGSRAPPSALSRSHPR-----TGSV-QTSP 1156

Db 1242 SLKSYSEVSECNKQMGSMRSPVAPLEGICRALPRGSPHSDKERTVLSGSIWQTP 1301
QY 1157 SSTGTGRGRGLPOLPRKGLTERSAMDIERNRQMKLNKYQVAGSDRLRQDYHSKR 1216
Db 1302 RATTSEEDGLKTP-----KQIRRESPIT-RAFGAIF 1333
QY 1217 SCMPHRCADVSTKSSSDVDVSAVSTSSASRSTSYMSVSEPRGNRKISVFTS 1276
Db 1334 KG-KPYDSITTIKENG-----RSIHEIPRQDILLQESKRTPREVOSTRIIEGSIQGP 1387
QY 1277 -KMONRGVSGKN-----LTSTISIGDWCSL-----KNDGSDTAVAL 1318
Db 1388 IKFNN-----SGOSAIKHNVSLITYPEKLSKGMPLLEVPENIKVGENGYEDVAGE- 1442
QY 1319 GTSCKRRSSIGAKNVAIVGSLRKSRSASQSGGKKTLSVQSRSETGA----- 1372
Db 1443 -----TVRSRHTSVVS---SGPSVLSSTLHAKQAQISPGIYD 1478
QY 1373 -----VEMRNMTROA-----SNESTGSMNSYSSEGNLIFPGV 1406
Db 1479 TSARPTVSYQNTMSRSGSPMNRSTDVITIPRKSTNHEKSTLTPTQRESIAPKSPVGV 1538
QY 1407 -----RLASDQSPDFLDGLPAQLVGRQTLATP--- 1435
Db 1539 DPVYSHSPDPHHRGSTAGEVYWSHLPTQLDPAMPFHRALDPAAMAAYLFQRLSPTPGY 1598
QY 1436 -----AMGDLYGVAMD-----KKGLEVEITRARGLYVKGSKTLPAPY 1474
Db 1599 SQYLYAMENRQITLNDYITSOQMQVNLRPVARGL--SPREOPLGLPY 1646

RESULT 11
NCR_MOUSE STANDARD: PRT: 1453 AA.

AC P30415; 01-Apr-1993 (Rel. 25, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NK-TUMOR RECOGNITION PROTEIN (NATURAL-KILLER CELLS CYCLOPHILIN-RELATED PROTEIN) (NK-TR PROTEIN).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93133824; PubMed-8421688;
RA Anderson S.R., Gallinger S., Roder J., Frey J., Young H.A.,
RT "A cyclophilin-related protein involved in the function of natural killer cells";
RT Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
CC -!- FUNCTION: COMPONENT OF A PUTATIVE TUMOR-RECOGNITION COMPLEX.
CC -!- INVOLVED IN THE FUNCTION OF NK CELLS.
CC -!- SIMILARITY: CONTAINS A CYCLOPHILIN-LIKE PPIASE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC
DR EMBL: L04289; AAA37500.1; ALT_TERM.
DR HSSP: Q27450; 1A58.
DR MGD: MGI:97346; Nkr.
DR InterPro: IPR002130; .
DR Pfam: PF00160; pro_isomerase; 1.
DR PROSITE: PS00170; GSA_PPIASE_1; 1.
DR PROSITE: PS50072; GSA_PPIASE_2; 1.
DR Cyclosporin; Isomerase; Rotamase; Repeat; Transmembrane.
FT DOMAIN 1 176 PPIASE, CYCLOPHILIN-TYPE.

FT DOMAIN 222 241 ARG/LYS-RICH (BASIC).
FT DOMAIN 422 459 ARG/LYS-RICH (BASIC).
FT DOMAIN 964 1003 ARG/LYS-RICH (BASIC).
FT DOMAIN 198 273 ARG/SER-RICH.
FT DOMAIN 468 565 ARG/SER-RICH.
FT DOMAIN 658 812 ARG/SER-RICH.
FT DOMAIN 1303 1453 ARG-SER TANDEM REPEAT-RICH.
SO SEQUENCE 1453 AA; 163439 MW; DF1173FF814B283E CRC64;
Query Match 3.0%; Score 248.5; DB 1; Length 1453;
Best Local Similarity 18.0%; Pred. No. 0.00014;
Matches 295; Conservative 185; Mismatches 570; Indels 591; Gaps 62;
QY 14 TPASQPPPEPMDLSHLEEKILLAVMDROK-----EE 51
Db 158 TDAASRPYADVRYIDCGVLA---TKLTQDYEFKKRRKPTGSEGDSSRSSSSSESSES 214
QY 52 EKESQVYKIKEENKAOPTQWFPFSGITELVNNVLOPOQKQNEKEPQTKLHQEFMYKEQ 111
Db 215 EVEEETIR-RRHKRRRK-----YHAKKRRKREMSSEPRKRTVSPGYSER 262
QY 112 VKMGEESQOQOEGKDPAPCGICHTKTFADGCGHNSYCOTKFCARCGGVLSRKNVM 171
Db 263 -SVNNEKRSVDSNTRKRPV----- 288
QY 172 WVCN---LCRQOEILTKSGAMFYNSGNTLQOPDQKVPRLRMEAPQEKAKLHQEPQ 228
Db 289 PVENRFLRLRDMFAIT-----VEPEONIP-----DVAPVSDQK----- 323
QY 229 FQGAPGDLVPAVER-GRAGLTROTDTKNGSGVYKHQIASDMPDRKRSVSPVRDONRY 287
Db 324 -----PSVSKSR-----KIKGCTIRYH-----TPRRSHSESKDDSE 360
QY 288 ---EQSEREDYSQVPSDGTMPSPSDYADRSCQREPQ-----FYEEFGH 330
Db 361 TPWHKKEEMQRLRAYRPPSGCKWMSKGDLPSCSSRWMEBSLSQSRWSWNGYSDIST 420
QY 331 LNRDNRGRGRHSHKEYIVDEDEVESDEYEROREEYQARYSD-----PVL-AR 381
Db 421 ARHSDGHHKHKRKKK-----KHKKAKQKQHCRRHROTCKRRTVMPDLRPSR 469
QY 382 YP---VYRQYEEQMRHAESYRARHERH---SDVSIANAELDESRIISLRMDRPSQ 434
Db 470 SPTHRMKSQCYRRR-RBRASSSSSHSKRDWMSKSDODDGSASHSSDYRSKSHSD 528
QY 435 RSVSERRAMENORSYMERTRERAGOSYTPQRTS----- 469
Db 529 SRGSSSRRAVSKSSSRSLNRSKSSSSSRSGCPRTSISPKKPAQLSENKPVTEPLRPSV 588
QY 470 -----NHSPTPRRSPILPD-RPDMRAQSLR-KOHLDPSSA-- 505
Db 589 QNGNVLVQPAENIPIYPLSDSPPRSMKPGQKRWKPSYERIDEMAKKTHLLPVQSTY 648
QY 506 ---VRKT-----KREKME-----TMR 519
Db 649 SLTIKATVSSSYHKKREKPESDGSAVSKYSDRSSGSSGSKSSRSRSTYTSR 708
QY 520 NDSISSQGSVPRPPRPPIKSKGKMKQVSLSS-----EELAS 561
Db 709 SRSLPTSRSLR-RSPSSSRSHSPNKYSQSHSRSSSYSVSSDGRAMFRNRRKKSYS 767
QY 562 TPETSSDDVLESESVSEKSDQKGRKKTSEQVLSDSNTRSEKQKRMYYG 614
Db 768 HKRHSRSEKTLHSHKYVRGKSSR-HRKYSNSGSLDYSQSHVQVSAPEKEQG 826
QY 615 -----GHSLEEDY---EMSEPOLKSDGVDTCSSTLNEHSHSDKHVYWPQSKDG 662
Db 827 KVELNDKQKGRGKRPKPEWECPRSKEN-----SEDS 862
QY 663 DRLIGRILLNKLKDGSVPRDSGAMLGKVVGGKMTESGRLCAFTYKKSGLADTVGHL 722
Db 863 -----RDDSVSKGKNCAGSKWDSSENSSEODVTKSRKSD-----P 896

```

OY 723 RPDGEVLEWNGRLLOGATFEFYNNIIIESKPPROYELVSRPISGD---IPRIIPSTHAOL 779
DB 897 RRGSEKEGEA---SSDSSEVGOSHITAKPPAKPTSTFLLPESDCAKMSRRQSSASES 953
OY 780 ESSSSFEQKMDRPSISVTSMPSPMLRDVDFOLSGOLSIKLMFDKVGHQLVTLIGAK 839
DB 954 ESSCENL-----GNINGEPQ-----K 969
OY 840 DLPSRDEGRPNRPVYKIFLPDRSDKNKRTKVTKTLPEKMNQTFIYSPVHRREFRERM 899
DB 970 QKHSKDLKGDH---TKRAREKSKAKDKKHKAKPKRQAFHWDP----- 1010
OY 960 LEITLMDQARVAREESEFELIETLLELTALLDDEPHMYKLOTHDVASLPLPRSPYLLPRR 959
DB 1011 -----PLEFG--DDE-----EE 1020
OY 960 OLHGESPTRRLORSKRISDSEVSDYDCE--DGVGVSDYRHNGRDLQSLVPEQ--V 1014
DB 1021 EKNKGQVTDQPEKRIHVSEK-----CEAVKDGIPNVYEKTCDEGSSPSKPKKGTLEDDPL 1074
OY 1015 MSSNCSPEGSGPHRVYVIGRTSRWSPPAPPOANNVEG-----HRG 1055
DB 1075 AEGGH-DPSSCAPLKEVEDNTASSPPSA---QHLEHHPGGEGEDVLOTDDMETCTPDRT 1130
OY 1056 TRATGHYNTISRMDRRVMDHYSDRD--CEAADROPYHRSRSTEQRPILLETTTSSR 1114
DB 1131 SPRAG--EYVSPPLANIRLDSPEVNIIPEODECMA-----HPRAGQESSMSFKTLGE 1182
OY 1115 SSERPTNLMRSPSLMTGSAAPPSPA-----LSRSHRTG---SVQTSPTSGTGGRG 1166
DB 1183 SGVKQSSSVTSVTSPEVSGKEGAEKQSMNLTKWKPLQGGNLVSTATTSSLDVKAL 1242
OY 1167 ROLPOLPEPKTLERSAMDIERN-----ROMKLN---KYKOVAGSDRLLEODY 1211
DB 1243 STYPEVKKPOL---REIKSKNKVRRGSLFDEVKRTARLRARRRNDQSSDDQTPSRDG 1298
OY 1212 HSKYRSGMDPHRGADIVYTSKSSDSDVSAVSTSSAPRSSSTSYMSVQSEPRGNKI 1271
DB 1299 DSGRS---PHRSRSKSETSRHRTRS--VSYSHRSRSRSTSYRSRSR----- 1348
OY 1272 SVTTSKMNQMGVSGKNLTKSTISGDMCSLEKNDGSDSTAVAGLTSGKKRRSICA 1331
DB 1349 -----SDMTSRGTRSRSS-----YSFHSRTRSSSR 1378
OY 1332 KNAIYGLSRKSRASQLOTEGGGKKLRSTVQBSTETGLAVEMKMMNTROASRSTGCS 1391
DB 1379 SRSSSYLHRSRSTYDYSSRSRSRSGRSDSYHRG-----RSYNRRSRSGR 1428
OY 1392 MNTSYSEGNLIFGVRLASDS 1412
DB 1429 SYGSDSESDRSYSHRSPSES 1449

RESULT 12
NCRC_HUMAN STANDARD: PRT: 1462 AA.
AC P30414;
DT 01-APR-1993 (Rel. 25, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NK-TUMOR RECOGNITION PROTEIN (NATURAL-KILLER CELLS CYCLOPHILIN-
DE RELATED PROTEIN) (NK-TR PROTEIN).
GN NKTR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93133824; PubMed=8421688;
RA Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A.,

```

```

RA Ortaldo J.R.;
RT "A cyclophilin-related protein involved in the function of natural
RT killer cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
RN [2]
RP REVISIONS.
RA Anderson S.K.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: COMPONENT OF A PUTATIVE TUMOR-RECOGNITION COMPLEX.
CC -!- INVOLVED IN THE FUNCTION OF NK CELLS.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE VIA ITS N-TERMINUS.
CC -!- SIMILARITY: CONTAINS A CYCLOPHILIN-LIKE PPIASE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL; L04288; AAA35734.2; .
DR EMBL; AF184110; AAD56402.1; .
DR PIR; A47328; A47328.
DR HSSP; Q27450; 1A58.
DR MIM; 161565; .
DR InterPro; IPR002130; .
DR Pfam; PF00160; Pro_Isoomerase; 1.
DR PRINTS; PR00153; CSA_PPIASE.1.
DR PROSITE; PS00170; CSA_PPIASE_1.
DR PROSITE; PS0072; CSA_PPIASE_2; 1.
KW Cyclosporin; Isomerase; Rotamase; Repeat; Transmembrane.
FT DOMAIN 1 176 ARG/LYS-RICH (BASIC).
FT DOMAIN 219 240 ARG/LYS-RICH (BASIC).
FT DOMAIN 421 457 ARG/LYS-RICH (BASIC).
FT DOMAIN 970 1010 ARG/LYS-RICH (BASIC).
FT DOMAIN 194 244 ARG/SER-RICH.
FT DOMAIN 466 574 ARG/SER-RICH.
FT DOMAIN 664 814 ARG/SER-RICH.
FT DOMAIN 1311 1348 ARG-SER TANDEM REPEAT-RICH.
FT SEQUENCE 1462 AA; 165676 MW; D98A1147763EF527 CRC64;
SQ

Query Match 2.9%; Score 244; DB 1; Length 1462;
Best Local Similarity 19.5%; Pred. No. 0.00023;
Matches 268; Conservative 169; Mismatches 501; Indels 434; Gaps 62;

OY 10 RPAETPAASQPPQPEKPLSLTEERKILLAVMDROKKEEKEGVLTKEENKAQPT 69
DB 299 RMPVVTAE---PEKIPDVAPTVSDQK-----PSVSKSGRIKRGRT 338
OY 70 QWPFSGITELVNNVLOPOQOPNEKE---PQTKLHQPEMYKEQYKMGESQOOEO 125
DB 339 -----IKYHTPPRRSCSESDDDDSSETPH---WKEEMQRLRAYRPPSGEK 382
OY 126 --KGDAPTGICHTKYPADCGCHNCSTCQTKFCARCGGRVLSRKNKMWYCNLCRKQEI 183
DB 383 WSKGCD-----KLSDCCSR-----W-----DERSL 402
OY 184 LITSGAMFYN-----SGSNLLOOPDOKVPRLGRNEAPQEKAKLHEOPQOG- 231
DB 403 SQRSSWSYNGYISDLSTARSHGHHKRRKEKKYK---HKKKGKKKHCRHNRKQTKKRI 459
OY 232 -APGDLSPAVENKGRAGLRTQDTIKNGSVK--HQIASD--MPSDRKSPSVSRQNRX 287
DB 460 LITPSLIESSKSTTRKMSQCDRERRSSSLSHSSKRDWMSDKDVQSSLTHTSSRDSY 519
OY 288 EQSEERPDYQYVPSDQTMRSPEVDYADRQQRPEQFEEEGHNLVYDSNRGHRNSKEY 347
DB 520 RSKSHSQSYSGSSRSRTASKS--SSHSRSRSKS-----SSSKSGHR--KRA 563
OY 348 IVDEDEVSREDEYRORREERYQARYSDPMIARYPV-----KROP----- 388

```

```

Db 564 SKSPRKATASQLESENKPVKTEPLRATMAQNNENVVQPVVAENIPVILPDSPPPSRKMKPCQ 623
OY 389 -----VEEQRIHAEEV-----SARHERHRSVDLSLMAELEDSE 422
Db 624 KPMKPSERIOEMKKTTHLLPIQSTYSLANKETGSSSYKRE-----KNSEDSQST 677
OY 423 ISLRMDRPS-----RQSVSERRAAMENORSYMERETREAOQSSYPO-----466
Db 678 YSKY-SDRSSSSSPRSRRSR-----SRYSRYTTRSLASSHSRSPSSRSHSN 730
OY 467 RTSNHSPTPPRSPPLPRDPMKRA-----DSLKQKH-----ID 501
Db 731 KYSDSQSSSSSYTSSISDDGRRAKRLRSSGKKNVSHKSSSEKTLHSKYVKGAD 790
OY 502 PSSAVRKTREKMETMLNDSLSDOSSESVPRPPRPHSKKKGKMKROYLSSEELAS 561
Db 791 RSSCVRKSESSESSL-----DYSSDSESSVQATQSAQEKQSG-----QMERHNNKEKNR 842
OY 562 TPPEYTSQDVELESESEVSEKGDQKRRKTSQGVLSNTRSEROKKRMYYGSHLEED 621
Db 843 G-----EKKSKSEBECPHSKKRTLKEN--LSDLRLNGSKPKRKNVAGS-----883
OY 622 LEMSPQTKDGVDTCSSTTLNEHSHDKHPYTWQPSKQGRLLGRILL-----NKRLK 676
Db 884 -KMPSESENSESDV-----TKNSKNDSPSSDKE--EGEATSDSESESEIHKYKPTTKSST 937
OY 677 DGSVPRDSGAMLGKLV-----VGKMTESGRLCAPITVYKKGSLADTV-GHLRPGDEVY 729
Db 938 NTSLPDDNGAMKSSKQRTSTSDSESCSSENNRCKPKQHKHGSKNLKRHTKKYKEKI 997
OY 730 E-----WNGRLQAGTFEEVYN--ILLESKPEPOVELVVRPIGDIP 769
Db 998 KKKDKKKHAKRKAQAFIMQPLERGESEEDIDKQVQSEKKEKVSE-----NNE 1049
OY 770 RIPDSTHAOLESSSSFSQKMDRPSISYTSMPSPGMLRDVQPLSGQSLIKIMEDKVG 829
Db 1050 TIKDNI--LKTEKSEEDLSGKHDTVYSS-----DLQFTKDSKLSISPTALMT 1098
OY 830 QLIYVIL-----GAKDLPREDGRPRNPVYKIFYLPDRSDKKRRKTYKTKTL 877
Db 1099 EENVACLONIQHVESVPRNGEDVLOTDDNME-----ICTPDRSSPAK-----VEET- 1145
OY 878 EPKWNQTFISVHRRERFERMLITL-WDQAR-----VREE--ESEFLGET-----921
Db 1146 -----SPLGNARLDTPDINIVLKDMATEHQAELVNAQESSMSKVLGEVKGD 1195
OY 922 -LIELETA-----LLDDEPHYKIQTHDVSLP-----948
Db 1196 SSSASIASAGESTGKKEVAEKSQIMLIDKK--WKPLQGVGNLAAPMAATSSAIVEKVLTT 1253
OY 949 LPRSPYLPRLQHGSESPTR-----RLORSKRISDSEVSDYDCEDGVGVVSDX 996
Db 1254 VPEMKPQGLRIETKKNKVPKPSLFEVAKTARLNRRPNQESS-----SDE 1300
OY 997 RHNGDLOSSTLVEQVWSSNHCSPSGSPHRAVDVIGRTSMSPAPPORNVEQCHRG 1056
Db 1301 QTPSRDDDSQSSS-PSRBSRK--SETKSRHRTSRVSYSHSRSRSTSYSTRSYSTR 1356
OY 1057 RATHGN--TISRMDRHRVMDHYSSDRDRCCEADROPYHRSRSTEOPLERTTTRS 1113
Db 1357 RSRGWSYSGKRTSRSSSYSTRSYSHRTSRSRSS-RSSSYSPHRSRSTYTDYSYSTRSR 1415
OY 1114 RSSEPDNINIMSPSLMTGRSAPPSPALSRSH-----PRGVSQTSST 1159
Db 1416 R-SQKSD-----SYHGRSYNRRSRSCRSYGSDESDESRYSIHRRSPSES 1458

```

```

DT 01-OCT-1994 (Rel. 30, last annotation update)
DE MICROTUBULE-ASSOCIATED PROTEIN 2 (MAP2B) [CONTAINS: MAP2C].
GN MAP2.
OS Rattus norvegicus (Rat.).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN (1)
RP SEQUENCE OF 1-1694 AND 1726-1861 FROM N.A.
RC STRAIN-WISTAR; TISSUE-Brain;
RX MEDLINE=90251471; PubMed=2339070;
RA Kindler S., Schwanke B., Schulz B., Garner C.C.;
RT "Complete cDNA sequence encoding rat high and low molecular weight
RT MAP2."
RL Nucleic Acids Res. 18:2822-2822(1990).
RN (2)
RP SEQUENCE OF 1-1694 AND 1726-1861 FROM N.A.
RC STRAIN-WISTAR; TISSUE-Brain;
RX MEDLINE=91060576; PubMed=2174050;
RA Kindler S., Schulz B., Goedert M., Garner C.C.;
RT "Molecular structure of microtubule-associated protein 2b and 2c from
RT rat brain."
RL J. Biol. Chem. 265:19679-19684(1990).
RN (3)
RP SEQUENCE OF 1-151; 1515-1694 AND 1726-1861 FROM N.A.
RX MEDLINE=90221819; PubMed=2326166;
RA Doll T., Papadimitrakopoulou A., Matus A.;
RT "Nucleotide and amino acid sequences of embryonic rat MAP2c."
RL Nucleic Acids Res. 18:361-361(1990).
RN (4)
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=89365159; PubMed=2770869;
RA Papadimitrakopoulou A., Doll T., Tucker R.P., Garner C.C., Matus A.;
RT "Embryonic MAP2 lacks the cross-linking stiearm sequences and
RT dendritic targeting signal of adult MAP2."
RL Nature 340:650-652(1989).
RN
RP SEQUENCE OF 1695-1725 FROM N.A.
RX MEDLINE=94110302; PubMed=8282767;
RA Doll T., Melchiorner M., Riederer B.M., Honegger P., Matus A.;
RT "An isoform of microtubule-associated protein 2 (MAP2) containing
RT four repeats of the tubulin-binding motif."
RL J. Cell Sci. 106:633-640(1993).
CC
CC -1- FUNCTION: THE EXACT FUNCTION OF MAP2 IS UNKNOWN BUT MAPS MAY
CC STABILIZE THE MICROTUBULES AGAINST DEPOLYMERIZATION. THEY ALSO
CC SEEM TO HAVE A STIFFENING EFFECT ON MICROTUBULES.
CC -1- ALTERNATIVE PRODUCTS: VARIOUS FORMS OF MAP2 ARE PRODUCED BY
CC ALTERNATIVE SPLICING OF THE SAME GENE. MAP2C, THE LOW MOLECULAR
CC FORM OF MAP2, LACKS THE CENTRAL DOMAIN OF MAP2A/B.
CC -1- DEVELOPMENTAL STAGE: MAP2C IS EXPRESSED DURING EMBRYONIC BRAIN
CC DEVELOPMENT AND UNTIL POSTNATAL DAY 10. MAP2B IS EXPRESSED
CC THROUGHOUT BRAIN DEVELOPMENT.
CC -1- SIMILARITY: CONTAINS 3 OR 4 TAU/MAP REPEATS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch.)
CC
CC EMBL: X51842; CA36135.1; -
CC EMBL: X17682; CA35667.1; -
CC EMBL: X71487; CA350588.1; -
CC PIR: S07887; S07887.
CC PIR: S10003; S10003.
CC PIR: A37981; A37981.
CC InterPro: IPR001084; -
CC Pfam: PF00418; tubulin-binding; 4.
CC PROSITE: PS00229; TAU_MAP_1; 3.
CC Microtubules; Repeat; Alternative splicing; Calmodulin-binding.
CC KW DOMAIN 1454 1474 CALMODULIN-BINDING (POTENTIAL).
FT

```

```

FT REPEAT 1664 1694 TAU/MAP MOTIF.
FT REPEAT 1695 1725 TAU/MAP MOTIF.
FT REPEAT 1726 1756 TAU/MAP MOTIF.
FT REPEAT 1757 1788 TAU/MAP MOTIF.
FT VARSPLIC 152 1514 MISSING (IN ISOFORM MAP2C).
FT VARSPLIC 1695 1725 MISSING (IN ISOFORM WITH 3 TAU/MAP
SQ REPEATS).
SEQUENCE 1861 AA: 202409 MW: 42DCFI16D1EF54E CRC64:

Query Match 2.9%; Score 242; DB 1; Length 1861:
Best Local Similarity 19.2%; Pred. No. 0.00038;
Matches 366; Conservative 238; Mismatches 655; Indels 646; Gaps 90;

8 GRDPAIPASQ-----PPPOFEMDLSHTEBEKKILAVMDROKKEBEKQSVL- 58
D 117 KDQPAALPLAAEETVNLPPSPSPASBOTALEED-----LITASKMEFPEOQKLPSSFA 172
Q 59 ---KIKEHKAQ--PTQMPFSGITELVNNVLOPOOKO-----PNEKEPQTKHQEFEM 107
D 173 EPLDKETEFEKMQSRGDEFEHVALVPQDTSKTPQDKKODPOMEKSPASPFQTFGT 232
Q 108 YKEOYKKMGESQOQOEOKGAPTCGIC-----HKTGFADGCGHNSYCOTKFCARC 159
D 223 NLEDIKQITEPSI-----TVPISIGLSAEPLAPKQKQWFL- 268
Q 160 GGRVSLRNNKVMWVGNIC-----RKQOEILTKSGAMF-----YNSGSNT 198
D 269 ---MVESEKDEM--GLAAPISPGILTPRKENDVLDPIMWEGKOFDPSWPSFHHGGSFT 323
Q 199 L---OOPQKVRGRL-----NEAPQEKKA-KLHEQ 226
D 324 LPLDYKIDRVEGSGPFAVPFQOSDKMSLDPTSGANSKSSKDEEPQKADKADVADY 383
Q 227 POPQCAP--GDLSPVAVE-----KGAHGLTROPTIKNS----- 259
D 384 PVSEATTVLGDVHSPAVEGVEGENISGEKGTDOEKETSTPSVQEPILTEPQTKLE 443
Q 260 -----GVKHQIASDMP-----DKRSPSVSRDONRREYQSEF--RED-----YS 297
D 444 EHSKVSIEIEVAKEESLKLKDDKAGVIGTSTQSFSEKQOQOETIEALKQDSFISL 503
Q 298 QYVPSDGTW-----PRSPS-----DYARR-----SOREPOFY 325
D 504 EQAVTAAAMTKTLEKYTSPEAVSEKREITQGLFEEDIDASKLEGASATAVALEMVFY 563
Q 326 EE--PGLHNYRDSNRGRHRSKEYIVDEDEVESRDEYERORREBYQARYSPDNLARYPV 384
D 564 EDSKSGSKYIETSA-----LKEDVTRSTGLGS--DYELLS-----DSRGNAQESLDTVSP 611
Q 385 KPOPYEOMRTHA-EVSRANHERHRSDVLANAELEDSRISLILMDRPSRQSRVSEKRAA 443
D 612 KNOODEKELLAKASQSPSPAHAGYS--TLAOSYTSHPSEL-----PEEPSSPOERHFT 664
Q 444 ME-----NORSYSMERTREAGQSSYPORTSNHSPPTPRKSPJPLDRPDMR 490
D 665 IDPKVYGEKRDLSKKNDDTLISLIGGSAIEQRMSINLPMSCDSTAL----- 717
Q 491 ADSLRKOHLDPPSSAVKTKREKMTMLRNDLSLSDQSESVRP-----PPRPHSK 542
D 718 GFNGRGHDLSPLAS-----DILNTSGMDEGDDYLPPTTPAVERKIPCFPIESK 767
Q 543 K-----GCKMROVSISSSEELASPEYTS-----CDVLELESSEVSE 580
D 768 FEEDKTEQAKVTGGQTVVE--TSSSPPAKEYKNGIYMAADLEBMDLAGTRSLAVY 826
Q 581 KGDSQKGRKRT--SEQGVLSNTRESEKOKKMYGSHLEEDLEWSEPOIKDSGVDTCS 639
D 827 SADAEVARRKSPSAVVAEST-----GLPVAVD--SQPVKPDSDLEDMGY 872
Q 640 TTLNEHSHSDKHPTW--QPSKQDRLIGRIILNKRILKSGSPRDSGAMGLKVVGGKMT 698
D 873 CVFN-----KYTVLPSPVODSENLISGSGSYEGTDKVRDLATLSTLIEV--KLA 923

```

```

QY 699 ESGRLCAFTIKVKKG---SLADTVGHLRGDEYLEMNGRL--LO----- 737
D 924 AAGRVDEFTEAEKSPSSADKSGLSREPDQDRKANDKLDIYLEKSEENVDSKEHAKES 983
QY 738 -----GATEEYVNNIL-----ESKPEPQVELVSRIGDIPRIPISTHQAQ--L 779
D 984 EEYGDVLEFLGLGVITYEQISAKELLITKETAPERAKEGLSSVP--EVAEVEYTTKADQGL 1041
QY 780 ESSSSSFESQKMDRPSISVTSWSPGMLRDYPOFLSQSLIKLMEFDYVGHQILYILGAK 839
D 1042 DVAKKDDQSPLD-IKVSDFGQMASGMSVDAGK-----TLEKFE--VDOQLI-----SS 1089
QY 840 DLPSREDGRNRNVKTYFLPDRSDKKRKRTKYTKLEPKMNQTFIYSPVNRREPERM 899
D 1090 EAPQETDS-----FWGIESHYKDGAK-----VSETEKEKV 1121
QY 900 LETIMDQARVREES--EFLQE--LILET-----ETALLDE----- 933
D 1122 AKPDVHQEAVDKRESEYSGEHESLIMESLKPDEGKETSPETSLIODEVALKLSVEIP 1181
QY 934 -----PHWYKLOTHDVSLP-----LPPSPYLPR--ROLHG 963
D 1182 CRRPVSEADSIDKAEVQMEFIQLPKEESTETPDIPAIPSDVYQPOPEANVSEBAEVRG 1241
QY 964 ESP-----TRILORS-----KRISDEVSDVCEGV 990
D 1242 FEETLEAGEVDKLLFRSDTLQIDLLVPGSREEFVETCPGBHKGIVESVYTIED--DFI 1299
QY 991 GVSVDYRRHNGDLOSSTLSVEQVWSSNHCSGSPGHVDV-----GRTSWSPAP- 1043
D 1300 TVYQTTTDEG--ELGSHSVFRAAPVQPEEERRRYPHDELEFLVMAAEQAEPDQSGPDAPA 1358
QY 1044 --PPOBNVQGRNGRATGHVNTISMDRHRVMDHYSSDRDCEADROPYHRSRSTEQ 1102
D 1359 TPERKEVPFSEKTYTDYDKDETID-----DSIMDASLWVDIQQDDRSITLITEQ 1409
QY 1103 RPLERTTTRSSSRPDTNLMRSPSLMTGSGAPSPALSHRPTSGVQTSPPSTPGT 1162
D 1410 LETIPKEERAKEARRPSLEKHKRKPFTGR-----GRI-----STP-- 1447
QY 1163 GRRGQLOLPKPGKLGERSAMDIERNROMKLNKYQVAGSDPRLQDYHYSKYRSGMDPH 1222
D 1448 -----ER-----REVAKKEP----- 1457
QY 1223 RGADVTSTKSDSDVSVAVSARTSSASRSTSYWSSVOSERPRGRKKSIVFTSKMQRQ 1282
D 1458 -----STVSRO--EYVRKKAAVYKKAELAKES-----EVOAHSP--SRKILLKPAIKYTRP 1503
QY 1283 KGVSGKNLTKSTISISGDMCSLEKNDSQSDIYVAGALGTSQKRR-----SSIGAKMVAIV 1337
D 1504 THLSQVCR--RKTTATSGESAQAPSAFROAKDKVTDGITKSPFKSSILPRPSIILPRRGVS 1562
QY 1338 GLSRKGRSASQLSQTEGGGKILRS--TVORSTETGLAVEMRNMWTRQASRSTGSMNYS 1396
D 1563 G--DRENSISLNSISASRTTSEPIIRAGKSGTSP-----TTPGSTAITQDTPSYS 1616
QY 1397 SEGNNLPGVRLASDSQSFSLDGLGPAQLVGRQIATLTPAMGDIQVAMMKGQLEVEII 1456
D 1617 SR-----TPGR-----PGTPSYRPTPSTPGSIL-----VPSKKAVALI 1650
QY 1457 RANGIYVKKPSKTLTPAVYKVYLNDNGVCIARK--KTKVARKTLEPLVQQLSFEESPOGR 1515
D 1651 R-----TPPKS--PATPKQRLINQPLPDLKNVKSXIG--STDMIKYO-----PKGG 1693
QY 1516 VLOII-----VWDYGRMDHKSFMGVAOILDELELSMW 1549
D 1694 QVRLNKKMDFSKVQSGSKMDINKHSGAGGNQVITKIKDLISHV 1738

```

RESULT 14
 TRHY_SHEEP STANDARD; PRT: 1549 AA.
 ID TRHY_SHEEP

AC P22793;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE TRICHOHYALIN.
 GN THH.
 OS Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 OC NCBI_TaxID:9940;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=93260018; PubMed=7684041;
 RA Fietz M.J., McLaughlin C.J., Campbell M.T., Rogers G.E.;
 RT "Analysis of the sheep trichohyalin gene: potential structural and
 calcium-binding roles of trichohyalin in the hair follicle.";
 RT J. Cell Biol. 121:855-865(1993).
 RN [2]
 RP SEQUENCE OF 1016-1549 FROM N.A.
 RC STRAIN=MERINO-DORSET HORN X BORDER LEICESTER; TISSUE=Wool follicles;
 RX MEDLINE=90130632; PubMed=2298812;
 RA Fietz M.J., Presland R.B., Rogers G.E.;
 RT "The cDNA-decoded amino acid sequence for trichohyalin, a
 differentiation marker in the hair follicle, contains a 23 amino acid
 repeat.";
 RT J. Cell Biol. 110:427-436(1990).
 CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
 IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
 INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
 LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
 ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
 WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
 ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
 ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
 DIFFERENTIATION.
 CC -1- SUBUNIT: HOMODIMER (PROBABLE). TWO FORMS MAY BE PRODUCED BY
 CC -1- ALTERNATIVE SPLICING OF THE SAME GENE.
 CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
 THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
 THE EPITHELIA OF THE TONGUE, HOOF AND RUMEN.
 CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
 CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
 ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
 OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
 ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
 THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
 CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
 THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
 CC DIFFERENT SPECIES.
 CC -1- PPM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
 CC PROBABLY CONVERTED TO CITRULINES BY PEPTIDYLARGININE DEIMINASE.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
 CC FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; Z18361; CA79165.1; -;
 DR EMBL; X51695; CA35992.1; -;
 DR PIR; A34209; A34209.
 DR PIR; S32633; S32633.
 DR PIR; A40691; A40691.
 DR HSSP; P02633; 31CB.
 DR InterPro; IPR001751; -;
 DR InterPro; IPR002048; -;

DR Pfam; PF01023; S_100; 1.
 DR Pfam; PF00036; efhand; 1.
 DR PROSITE; PS00018; EF_HAND; 1.
 DR PROSITE; PS00303; S100_CABP; FALSE_NEG.
 DR Repeat; Calcium-binding.
 KW DOMAIN 1 91
 FT CA_BIND 22 33
 FT CA_BIND 62 73
 FT DOMAIN 413 832
 FT REPEAT 413 448
 FT REPEAT 449 476
 FT REPEAT 477 504
 FT REPEAT 505 532
 FT REPEAT 533 560
 FT REPEAT 561 588
 FT REPEAT 589 616
 FT REPEAT 617 644
 FT REPEAT 645 678
 FT REPEAT 679 706
 FT REPEAT 707 742
 FT REPEAT 743 771
 FT REPEAT 772 796
 FT REPEAT 797 832
 FT DOMAIN 938 961
 FT REPEAT 962 985
 FT REPEAT 986 1021
 FT REPEAT 1022 1044
 FT REPEAT 1045 1067
 FT REPEAT 1068 1090
 FT REPEAT 1091 1121
 FT REPEAT 1122 1144
 FT REPEAT 1145 1167
 FT REPEAT 1168 1197
 FT REPEAT 1198 1227
 FT REPEAT 1228 1250
 FT REPEAT 1251 1273
 FT REPEAT 1274 1296
 FT REPEAT 1297 1319
 FT REPEAT 1320 1342
 FT REPEAT 1343 1368
 FT REPEAT 1369 1391
 FT REPEAT 1392 1416
 FT REPEAT 1417 1439
 FT REPEAT 1440 1461
 FT REPEAT 1462 1484
 FT REPEAT 1485 1507
 FT VARIANT 1145 1197
 FT VARIANT 1251 1273
 FT CONFLICT 1399 1399
 SO SEQUENCE 1549 AA; 201173 MW; E72FB9FE1326E54E CRC64;
 Query Match 2.98; Score 241; DB 1; Length 1549;
 Best local Similarity 17.7%; Pred. No. 0.00034;
 Matches 229; Conservativity 185; Mismatches 518; Indels 362; Gaps 43;
 QY 32 LTFEERKIIIAVMDRQKKEEESVLKIKKEHKAQ-----TQMFPSGITE 79
 DB 252 LLEDER-----EQLRRKEDEREDQQLQEQEATQEEISRGESRTSRGW-----Q 300
 QY 80 LVNVVLQPOOK---QPNKEPQTKLHQEFMYKQVKMGESQOQOQOKGAPTCGICH 136
 DB 301 LESAADARQKRVYSRPHRQEQOSRQEBELLEROEQOISEFVSLQDQGR-----Q 353
 QY 137 KTRFADCGHNCYS-----CQTKFCARGGVSLSNKNVMV-----CNLCRQKQETL 184
 DB 354 RLKQEQRYDQWWRQQLLEESQRRRYTLAKPAQREQVREBQLRLKEKRLQREKRQRE 413
 QY 185 TKSQAWFYNSGNTLQPDQKVPVPGRLRNEAPQER-----KAKLHQDPQPGAGDLS- 237
 DB 414 RQYREVELQREERLQREBQLOREERKRRRQREKQYLEKVELMEBQLOREERERRR 473

```

OY 238 -----VPAVEKGRAGLTRODTIKNGSSGVKHLASDMPSDRKSPEVSDONREYQS 290
DB 474 QERREKQYLEVEELREEBOLQOREKRRQREKROYLEKVV-ELQEEBOLQOREKRRQRE 531
OY 291 EEREDYQOYVPSDGTMPSPSDYADRRSQRBPOVE-----EPGLNTRDSNRGRHSHK 345
DB 532 KERO-YLEKVELQEEBOLQOREKRRQREKROYLEKVELQEEBOLQOREKRRQREK 590
OY 346 EYIVDEEDVESRDEYERQOREEEOQARTRSDPNLARYPVKPPYVEQOMTHAEVSARHE 405
DB 591 QYL---EKVELQEEBOLQOREKRRQREKROYLEK---VELOEEBQVQOREKRRQRE 645
OY 406 RRHSDVSLANAE-LEDSRISILRADRPSR-----QRSVSERRAA 443
DB 646 RQYLEKELQOREBOLQEEB-QLTRERERKRRQREKROYLEKVELQEEBOLQOREKRR- 703
OY 444 MENORSYSMERTREAGOGSSYTPORTSNHSPPTPRSPITLDRPDMRARSILRKHLLDS 503
DB 704 QERREKQY-LEK-EELOREBOLQOREK---LQREKRRQREKROYLEK 750
OY 504 SAVRTKREKMETMLRNDSLSDSQSESVAPPPPHKSKGGMQVSLSSSEELASTP 563
DB 751 LQOEDRLQOREQLREB-----REKR----- 772
OY 564 EYTSQDDVELESESVSEKGSQKRRKTSQGYLSDSN-----TRESQKRRMYG 614
DB 773 QYLEKVELQOREBOLQOREKRRQREKROYLEELREEBELHREKQOLQOREKRRR--- 829
OY 615 GHSLEEDLEMSRPOIKDSQVDTCCSTTLNDEHSHSKHPYTWQPSKDGDLIGLILNK- 673
DB 830 -QELRQOLEEELQRLDR-----KQRFDDDOH---QNEVANSVYVSHREKKE 874
OY 674 --RLKDGSVPRDSGAMIGLVGKMGTEGRLCAFTTKVKKGLADTVGHLRQDEVLW 731
DB 875 KSRQOLDMSVRESQFQDLRLQODEOEK-----REBQEW 910
OY 732 NGRLLOGATFEVYNNITLESKPPQYELVYSRPIGDIPTDTHADLESSSESOKM 791
DB 911 RSKOKRDSQFPA-----EOLLEFQOKETPRR 937
OY 792 DRRSIVTSPMSGMLRDVPOFLSGQLSIKLMFDKVGHOLIVTLLAKDLPREDGPRN 851
DB 938 DR-----KREBQOLKQOREKIRY-----LEDRKFR 967
OY 852 PVYKIYFLPDRSDKNKRTYVKTLEPKMNQFTYSPVHRRER-----ERMLLETL 904
DB 968 EEOQLRLREEOQLRQERDRKFRREELSRQERD-----RKFRREBOLQOREBOLR 1018
OY 905 WQDARRESESEFLGELLIEL-----ETALLDEPHWYKLTQHDVSSILPLRPS 953
DB 1019 QERDRKFRREBOLQOREBOLRQERDRKFRREBOLQOREBOLRQERD----- 1068
OY 954 PLYPRLOHGESPTRLORSKRISDSSEVSDYDCEDGVGVSDYRNGDLOSTLSVBO 1013
DB 1069 ---RKFRREBQOLRLLEBQOLRO-----ENKRFREBOLRERE 1106
OY 1014 VMSNHCSPSSPHRVNIVGTRBSMSPSPAPQARNVEOGHCRATGHNNTISRDNRHV 1073
DB 1107 QLRLOEGEPOLRQK-----RDRKFHEBOLQOREBQ-----LRQERDRK 1147
OY 1074 MDDVSSDRDDBCAADROPYHRSNSTBORPLLETTTRSSSE-----RPDTN 1122
DB 1148 FREBAQILKEHE-BQLRQNDKFRKFRREBOLQOREBOLRQERDRKFRREBQ 1206
OY 1123 LMRSMPSLMTGSRAPSPALSSHPRTGSVQVSPSTPGTGRGRGLPOLPPKGLTERSA 1182
DB 1207 LLOERBKL---RQERBOLRQERDRKFRHEBOLQ-----REBOLRQERDRKFR 1258
OY 1183 MDIERBNOMKLNKYKOVASDPRLEODYHNSKYR 1216
DB 1259 QLOERBOLRQERDRKFRREBOLQOREBOLR 1292

```

```

RESULT 15
ATRX_MOUSE ID ATRX_MOUSE STANDARD: PRT: 2476 AA.
AC 061687;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TRANSCRIPTIONAL REGULATOR ATRX (X-LINKED NUCLEAR PROTEIN)
DE (HETEROCHROMATIN PROTEIN 2) (HPI ALPHA-INTERACTING PROTEIN) (HPI-BP38
DE PROTEIN).
DE ATRX OR XNP OR HP1BP2.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98213653; PubMed=9545503;
RA Picketts D.J., Tasthan A.O., Higgs D.R., Gibbons R.J.;
RT "Comparison of the human and murine ATRX gene identifies highly
RT conserved, functionally important domains.";
Mamm. Genome 9:400-403(1998).
RN [2]
RP SEQUENCE OF 325-1176 FROM N.A.
RX MEDLINE=97133299; PubMed=8978696;
RA le Douarin B., Nielsen A.L., Garnier J.-M., Ichinose H.,
RA Jeanmougin F., Losson R., Chambon P.;
RT "A possible involvement of Tif1 alpha and Tif1 beta in the epigenetic
RT control of transcription by nuclear receptors.";
RL EMO J. 15:6701-6715(1996)
CC -1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PHF-FINGER DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF026032; AAC08741.1; -.
CC DR EMBL: X99643; CAA67962.1; -.
CC DR MGD; MGI:103067; xnp.
CC DR InterPro: IPR000330; -.
CC DR InterPro: IPR001650; -.
CC DR Pfam; PF00176; SNF2_N.1.
CC DR DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding.
CC KW DOMAIN 219 267
CC FT NP_BIND 1579 1586
CC FT SITE 1704 1707
CC FT DOMAIN 319 322
CC FT DOMAIN 735 738
CC FT DOMAIN 1001 1004
CC FT DOMAIN 1130 1135
CC FT DOMAIN 1182 1185
CC FT DOMAIN 1238 1245
CC FT DOMAIN 1484 1487
CC FT DOMAIN 1924 1931
CC FT DOMAIN 2205 2208
CC FT DOMAIN 2245 2248
CC FT DOMAIN 2403 2408
CC FT SEQUENCE 2476 AA; 278601 MW; 90A42B790FC4FF4C CRC64;

```

Query Match 2.9%; Score 239.5; DB 1; Length 2476;
 Best Local Similarity 19.0%; Pred. No. 0.00071;
 Matches 329; Conservative 231; Mismatches 616; Indels 553; Gaps 80;

```

QY 37 KRIIAVNDROKKEEBKESVLKIKEHKAOPTOM-----PESGITELVANNVLO--P 87
Db 240 KCIILRNIGR-----KELST--IMDEN-----NOMYCIQPEPLLDIVATACNSVFENLE 287
QY 88 OOKQONER-----EPQYTLHQO-----FEMYKPOVKKMGESQOQOQKQ----- 127
Db 288 OLOONKRRKIKVDSERTKSVCDOTSFKSPKSSSCNGEKKLEESCSCGVSTYSHAL 347
QY 128 DAPTCGICHTKTFADGCGH--NCSYCOTKFCARGCGRVSLRSNNVMWVNCV--CRKQOEIL 184
Db 348 SVFKEMIKTKTTLIETSMNSY--IKFLQAADNEMTS--AMLLCQLKSTKSVLDDI 403
QY 185 TKSQAMFYNSGNTLQOQD-----QKVPGLRNEAPQO-----KKAHLEQPOFOGAP 234
Db 404 KKAHLEEDLNEIOLADDVHKKEKNTKDLKSTDAKSETKLGKESYSYTEREFLKLA 463
QY 235 DLSVAV--EKRAGLTQODITKNGSVKHOIASDMPDRKSPSVSRQONRYEQSEE 292
Db 464 RSSVKAIDGEEQRAHKTSGE--HKSG-----RKGSGQYEPTNT 501
QY 293 REDYS-----QYVPSD-----GTMPSPSDY--ADRSOREPOFEYEPGHLNRYDSN 337
Db 502 PEDLDMDIVSVSSVPEDIFDLSIEMAYOSSADYOGDNGSGTEPELESVYLVNYSKD 561
QY 338 RGHHRHKEYIVDDEVEDERDEYERQOREEYQARYSD-----PNLARYPVKQPYE 390
Db 562 SGNKJRSK-----YTAKYRKELEFVKLTPLVSLNSPIKGVDCQ 598
QY 391 EOMRHAHYSRARHERRHSDVSLANAELEDSRISLMDRPSQORSYSEKRAAMENORST 450
Db 599 -----EVSOKNGKRSKSGVA-----RSSEKCRPREISISHENNVTI 634
QY 451 SMERTREAOGOSYPORTSNHSPRPSPPIPLDRPDMRADSLRQOHLDPSSAVAKTK 510
Db 635 LLEDS-----DLRSPRYKTTPLRQOTESNPAMNSNDESGNGMK 675
QY 511 REKMETMLRNDLSLSDQESVRRPPRPKSKGKMGROYST-----SSSEELASTPE-- 564
Db 676 KOKMGPIRKK--DKRNSADCATDNPKPHKVPKA--KQVIGDONSDSDEMLAVLKEAS 730
QY 565 ---YSCDVELESYSKSG-----DSOKGRKRTSEQVLSNSNRS-----ERQ 607
Db 731 OGHSSSSSTJDINEPOMNKGKTGKDNKGRKRNKSTSG--SDFTKKGKSTETSIISK 788
QY 608 KKRMYTGGHLEEDLEWSEPOIKDSGVDTCSSYTLNEHSHSKHPVTWQPSKDGRLI-- 666
Db 789 KRONTSSESNYSSELEREIKTMSRIGAKKSVPEKKEBDSDE-----KQKKVVD 840
QY 667 -GRILLNRKLDGVPYRDSGAMLGKLVGKMTESGRLCAFITKVKKGLADTVGHLRPG 725
Db 841 NGHERAKTTGGSSADDTGDTGGRQ--GGSCSIAG--GSIKVRSGV-----EFR-- 887
QY 726 DEVLEMNGLLOGATEEYVNTILESKPEPOVELVYSRPIGLIPRIPLDST----- 775
Db 888 -BMLCKPGVSSDGA-----EKPSYKEENVNSPEDKRVSKTKKTKHLRSRQSRK 935
QY 776 -----HAOLESSSSSFESOKMDRPSISVTSMPSPGMLRDVQFLSGOLSIKLF 824
Db 936 GKGSSSDGTDRPPKROSDSSSEGEKKQ-----SRQPGTGKAKAPDLGE----- 981
QY 825 DKVGHQIIVTILGAKDLPRDEGRPNPYVKIYFLPDRSD-----KNKRTKTVKKTLEP 879
Db 982 -----TLKREQWSSSDQTER-----LPEEEHIGPFSKGIKOSKTDTAGEK 1024
QY 880 KMNQTFIYSPVHRERREMLETTLMDQARVRESEFLEJLIELETA--LLDDEPHWYK 938
Db 1025 K-----GKKWKDKSCF-----KKEELSDVYKLPKGKSDSCSSSEDKTNR 1065
QY 939 LQTHVSSLPRLPRSPYL-----PRRLHGESPTRLQRSKRISDSVSDYDCEG 990
Db 1066 VSLREKKRSLPAKSGRPECSSDTEKSLKQCCDSTKRRKRIDLRRRN----- 1118
QY 991 GVSVDYRHNGRDLQSTSLVPEQVMSNHCSFGSPHRYVIGRTNSWSPAPPORAVE 1050

```

```

Db 1119 ---SSSKRNTKEVNSAS-----SSSDAGSSSDNKKOKKORF-----SAKKKTGNK 1162
QY 1051 QGHRGTRATGIVNTISNRDRHVRMDHYSDBDRDCEAADRPYHRNSTEQ--RPLLR 1108
Db 1163 EKKR-----NSLRATPKRKQYDITSSS--DIGDDONSAGEESSDEOKIPEVEN 1211
QY 1109 TTRTSRS--SERPDNTLMRSMPSLMTGRSAPPSP-----ALSRSHPRTGSV 1152
Db 1212 LVPLSHHGFQSSGSDDEALSKSVPATVDDDDNDNDENHIAKMLLEIKANLSDSDGSS 1271
QY 1153 QTSPPSTPGTGRGRQOLPOLPPKGLTERSAMDIEERNQOMKINKYQVAGSPRLBODYH 1212
Db 1272 DDEPDG--GKKRIGIKQSESP-----ADGELRRQDLAVQVNSSESDSDS--EESK 1320
QY 1213 SKYSGMDPH-----RGADIVTSTSSDSDVDY--SAVSRTS 1247
Db 1321 PRYHRLIRHLTLSDGSEGEKPTKPEHKEAKRRNRKVSSESDSDTDFQESGVSEEV 1380
QY 1248 SASRFSSTSYMSVOSERPGRGNKISV-----FTSKMONRQMGVSGKNLTGSTSISGDMC 1301
Db 1381 SESE-----DEQRPRTRSKAKKALEENORSYKQKKRRRIKYQEDSSSEKSHSE-- 1431
QY 1302 SLEKNDGSQDPAVGALGTSGKRRRSSIGAKWVAIVGLSRKRSASQLSQTEG--GKKLR 1360
Db 1432 --DKKEGDEEDE-----DEDEDEDENDDSKSPGKGRKIR 1466
QY 1361 STVQSTETGLAVEMRNMTRQASRESTDGSNYSSEBGNLFPVGRLASDQSFDFDG 1420
Db 1467 KILK--DDKLRTETONALKEEERRK-----RIARERERERKL-- 1502
QY 1421 LGPQOLVGRQTL--ATPAMGDIQVM--MDKGOLEVEIIRA--RGIIVKPGSKTLPAVY 1474
Db 1503 -----REVIEIEDASPTKPIITTKLVLDENETKEPLVGVHNMVAK----- 1544
QY 1475 VKVYLLDNGV-----CIAKKTIVARKTLEPLYQQLSFEESP-----QGRV 1516
Db 1545 LKPHQVD--GVQFMMDCCCESEVEKTK-----KSPGSGCIIAHCMGLGKT 1586
QY 1547 LQIIVMGDYGMDHKSFMGVAOILLDELELS-----NWVIGPKLP 1557
Db 1587 LQVV-----SFLHTV--LLCDKLDFTALVYVCLNTALNMWNEF 1623

```

Search completed: November 21, 2001, 16:10:18
 Job time: 239 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2001, 16:00:24 ; Search time 19.55 Seconds
(without alignments)
1830.191 Million cell updates/sec

Title: US-09-617-099b-1

Perfect score: 8285

Sequence: 1 MSAPLGPGRAPTRPAASQP.....TRRAGSSLSSTNGCPYSRS 1590

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRNUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------------|-------------------|
| 1 | 307 | 3.7 | 1898 | 1 US-08-056-200-94 | Sequence 94, Appl |
| 2 | 307 | 3.7 | 1898 | 2 US-08-800-644-94 | Sequence 94, Appl |
| 3 | 258.5 | 3.1 | 2842 | 1 US-07-741-940-7 | Sequence 7, Appl |
| 4 | 258.5 | 3.1 | 2842 | 1 US-08-289-548A-7 | Sequence 7, Appl |
| 5 | 258.5 | 3.1 | 2842 | 1 US-08-452-654-7 | Sequence 7, Appl |
| 6 | 258.5 | 3.1 | 2843 | 1 US-08-452-655B-2 | Sequence 7, Appl |
| 7 | 258.5 | 3.1 | 2843 | 1 US-08-452-655B-7 | Sequence 7, Appl |
| 8 | 258.5 | 3.1 | 2843 | 3 US-08-450-582-2 | Sequence 2, Appl |
| 9 | 258.5 | 3.1 | 2843 | 3 US-08-450-582-7 | Sequence 7, Appl |
| 10 | 258.5 | 3.1 | 2973 | 2 US-08-821-355A-7 | Sequence 7, Appl |
| 11 | 258.5 | 3.1 | 2973 | 2 US-09-003-687A-7 | Sequence 7, Appl |
| 12 | 258.5 | 3.1 | 2973 | 4 US-09-136-605-7 | Sequence 7, Appl |
| 13 | 250.5 | 3.0 | 2843 | 1 US-07-741-940-2 | Sequence 2, Appl |
| 14 | 250.5 | 3.0 | 2843 | 1 US-08-289-548A-2 | Sequence 2, Appl |
| 15 | 250.5 | 3.0 | 2843 | 1 US-08-452-654-2 | Sequence 2, Appl |
| 16 | 250.5 | 3.0 | 2843 | 2 US-08-370-235A-2 | Sequence 2, Appl |
| 17 | 225.5 | 2.7 | 1780 | 1 US-08-769-309A-5 | Sequence 5, Appl |
| 18 | 225.5 | 2.7 | 1780 | 3 US-08-994-570-5 | Sequence 5, Appl |
| 19 | 222 | 2.7 | 2101 | 1 US-08-466-390-4 | Sequence 4, Appl |
| 20 | 222 | 2.7 | 2101 | 1 US-08-467-781-4 | Sequence 4, Appl |
| 21 | 222 | 2.7 | 2101 | 1 US-08-470-950-4 | Sequence 4, Appl |
| 22 | 222 | 2.7 | 2101 | 2 US-08-483-924-4 | Sequence 4, Appl |
| 23 | 222 | 2.7 | 2101 | 4 US-09-452-284-1 | Sequence 1, Appl |
| 24 | 219 | 2.6 | 2101 | 1 US-08-195-487-4 | Sequence 4, Appl |
| 25 | 219 | 2.6 | 2101 | 5 PCT-US93-06160-4 | Sequence 2, Appl |
| 26 | 206 | 2.5 | 1805 | 1 US-07-853-913-2 | Sequence 2, Appl |
| 27 | 205 | 2.5 | 1162 | 2 US-08-728-323A-2 | Sequence 2, Appl |

| | | | | | |
|----|-------|-----|------|---------------------|-------------------|
| 28 | 200 | 2.4 | 927 | 3 US-08-895-601-6 | Sequence 6, Appl |
| 29 | 197.5 | 2.4 | 3669 | 4 US-08-061-376-5 | Sequence 5, Appl |
| 30 | 195 | 2.4 | 1068 | 1 US-08-396-479B-12 | Sequence 12, Appl |
| 31 | 195 | 2.4 | 1068 | 1 US-08-818-823-12 | Sequence 12, Appl |
| 32 | 193.5 | 2.3 | 1658 | 2 US-08-609-049A-13 | Sequence 13, Appl |
| 33 | 193.5 | 2.3 | 1658 | 4 US-09-170-996-13 | Sequence 13, Appl |
| 34 | 193.5 | 2.3 | 1726 | 2 US-08-609-049A-30 | Sequence 30, Appl |
| 35 | 193.5 | 2.3 | 1726 | 4 US-09-170-996-30 | Sequence 30, Appl |
| 36 | 190 | 2.3 | 1360 | 4 US-09-393-569-2 | Sequence 2, Appl |
| 37 | 188.5 | 2.3 | 1075 | 5 PCT-US94-07297-41 | Sequence 41, Appl |
| 38 | 185.5 | 2.2 | 1210 | 1 US-08-320-559-26 | Sequence 26, Appl |
| 39 | 185.5 | 2.2 | 1210 | 3 US-08-545-860D-26 | Sequence 26, Appl |
| 40 | 185.5 | 2.2 | 1210 | 5 PCT-US94-04496-25 | Sequence 25, Appl |
| 41 | 182.5 | 2.2 | 137 | 2 US-08-609-049A-15 | Sequence 15, Appl |
| 42 | 182.5 | 2.2 | 137 | 4 US-09-170-996-15 | Sequence 15, Appl |
| 43 | 182 | 2.2 | 1706 | 2 US-08-459-568-2 | Sequence 2, Appl |
| 44 | 182 | 2.2 | 1706 | 2 US-08-399-411-2 | Sequence 2, Appl |
| 45 | 182 | 2.2 | 1706 | 3 US-08-516-859A-2 | Sequence 2, Appl |

ALIGNMENTS

RESULT 1
US-08-056-200-94
; Sequence 94, Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Methods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/056,200
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054,001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-056-200-94

Query Match 3.7%; Score 307; DB 1; Length 1898;
Best Local Similarity 17.4%; Pred. No. 2.8e-14;
Matches 245; Conservative 213; Mismatches 514; Indels 434; Gaps 53;
QY 19 OPPPEMPPLSHUTEE--RKILAVMDQKKKEKESVLIKKEHKAPQTQWFPSSG 76

```

Db 264 EPQORELQOE-----EEQOLKLEERQELRREQEEOQOLRLRREQOLRRKQOE----- 312
QY 77 ITELANNVLOPOOKORNEKPEPTKILHOQFMYTEOYKAKKEEEOOQO---EOKGADPTCG 133
Db 313 -----ERRQOEERREOQOLRRQOEERREOQOLRRQOEERREOQOLRRQOE----- 358
QY 134 ICHKTAFADGCGHNSYCQTKFCARCGRVSLSKNKVMWVCNLCKOQELLTSGAMFYN 193
Db 359 -----RRQOQL----- 364
QY 194 SSSNTLOQPOQVPRGLRNEAPQOEKAKLHEQPOFQAGBDLSVPAVEKRAHGLTRQD 253
Db 365 -----RREOQEERREOQOLRRQOEERREOQOLRRQOQLRRQO-----QOLRREOQOLRRQ 412
QY 254 TIKNSGVNHOJASDMPDRKRSPSPSRQONKRYEGSEER-EDYSQVPSDGTMPSPSD 312
Db 413 QLRREOQLRREOQLRREOQLRREOQLRREOQLRREOQEHQKHEQERREOQLRQOE 472
QY 313 YAD-----RRSOREPOFYEEEPGH---LNYRDSNRGHRHSKEYIYDDEQVE 355
Db 473 RRDWLKREETERHEQERKQOLKQOEERRRRWLKLBEERREOQERREOQLRQOE 532
QY 356 SDE- YEROREEEXQARYRSDPNLARYPVQPYEEMRIHAESRARHERHSDVSLA 414
Db 533 RREOQLKREOEEERLOQRLRSEQOLRR---EQEERLEQLKREERKLEQERREOQLR--- 586
QY 415 NAELEDSRISLMDRPSRQOSYSEKRAMENORSTSMETRAQOSSTPQOTSNSHP 474
Db 587 KREOERROQLKREERROQLKREOERLEQR-LKREVERLEOERDERERLKRKEPE 645
QY 475 YRRSPPIPRDMRADSILRKOHLDPSAVALTKREKMETMLRNDSSDQSESRRP 534
Db 646 EERRH-----ELKSEQEEERH---EQLRREOQERREOQLKRE---EEERLEQR 690
QY 535 PPRPKKSGKGMROYSLSSSEBELA-----STPYTSCDDVELESSEVSSEKSGDSQK 588
Db 691 LKREHEER---RQOLAEEOEQARERIKSRIPKW---QWOLESEA-----DARQSK 737
QY 589 -RKTSEOGVLSDSNTRSEKOKRMYYGSHLEEDLEWSEPOIKDSQVDCSSTLMEHS 647
Db 738 VILLERPOAGRAAPQOEKKRR-----BSELQMOE-----EERA 772
QY 648 H-----SDKHEVTWQPSKGDRLIGRLLNRLKDGSVPRDSGAMLGKVVGKMTESG 701
Db 773 HROQOEERQRPFTWQWQAEKSEKSGRQRLSAR---PPLRQERQRLRAEERQOERQ 827
QY 702 RUCATITTVKKGSLADYVGHLPQGVLEWNGRLLQAGATFEVYNTILSKPEPOVELV 761
Db 828 PL-----PEEKEEORGR-----QRREKELOF 851
QY 762 SRPIDIPRIDPSTHAOLESSSSSFESOKMDRPSISVTPSPMGMLRDVPOLISGOLSK 821
Db 852 -----LFE-----EEOLOREKRAQOLOEEDDEGOEDERRROERQDQ 889
QY 822 LWFQVGHQILVTLIGAKDLPSEDEGRPNPYKTYFLPDRSDKNKRTKYAKTLEPKW 881
Db 890 KW-----RQWL-----EERKRRRHT-LYAKPALQOLKKEEQOLLOEEOEEO 931
QY 882 NOTFIYSVHREPEREMLLEITLMQAVRESESEFLGILILETALLDDEPHWY----- 937
Db 932 REE---RKRRRQOEEROYR---EEEOLOQOEEOQLRE---EREKRRQEREROYRKDK 981
QY 938 KLOTHIVASLPLRPSPLPRROLGESP--TRLORSKRISDSSEVDQDCEDGVGVSD 995
Db 982 KLOQKE-----EQLGEEPERKRRQERREKTYREE----- 1011
QY 996 YRHNGRDLOSSTLSVPEQVMSNMHSCSPHARDVIGTRTSM---PSAPPQORVNEG 1052
Db 1012 -----ELQOE---EEQLLRERER-----RRQWMEYOYRKDKDLQOEEOQL 1050
QY 1053 HRCGRATGHYNTISMGRHVRWMDHYSSDRDQCEADROPYHRSRSTEOPLLETTTR 1112

```

```

Db 1051 LREER-----EKRLREREROYREEELOQOEEOQLGERETRRROELFRQYR 1099
QY 1113 SRSEPRDPTNMRSPSLMTGRSAPPALSKRSHPRTSQVOTSPSPSTQGRGQOLPOL 1172
Db 1100 EEELOQOEEOQLRE-----EPKRRRQEREROQREEELOQOEEOQLLREERK 1147
QY 1173 PPKOTLEASMDIERNNQMLINKYQVAGSDPRLEQDYHSKYRSGWPH----- 1222
Db 1148 RRQOLEROYEEELQOKKRRQYR-----DEQORSDLKQWQEPKEKENVANDNKV 1198
QY 1223 --RGADTVSTKS-SDSDVSDVSAVSRTSSASRFSSTYSWYOSEPRGNRKISVFTSKMO 1279
Db 1199 YCKRENOQFQOLDSQYRD-----KQSQOQLQHLIGEQEERD-----EOE 1240
QY 1280 NRQVSGSKNLTSTSLSGDMCSLEKNDGSDQTAVALGTSKKRRSSIGAKMYAIVGL 1339
Db 1241 RRMQOANRHPPEE-----QLEREBQKA-----KRR----- 1268
QY 1340 SRKRSASQLSQTEGCGKKLASTYQR 1365
Db 1269 DRKSQOEKQLREERERRRQETDR 1294

```

RESULT 2

```

US-08-800-644-94
; Sequence 94, Application US/08800644
; Patent No. 5938752
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichothyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Methods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Knobe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,644
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/056,200
; FILING DATE: 30-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054,001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ. ID NO.: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-800-644-94

```

Query Match 3.7%; Score 307; DB 2; Length 1898;
 Best Local Similarity 17.4%; Pred. No. 2,8e-14;

Matches 245; Conservative 213; Mismatches 514; Indels 434; Gaps 53;

[illegible]

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

MOLECULE TYPE: protein
 ORIGINAL SOURCE: Homo sapiens
 IMMEDIATE SOURCE: APC
 CLONE: APC
 US-07-741-940-7

Query Match 3.1%; Score 258.5; DB 1; Length 2842;
 Best Local Similarity 17.9%; Pred. No. 2.7e-10;
 Matches 304; Conservative 238; Mismatches 618; Indels 535; Gaps 69;

30 SHLSEEEKILLAVNDQKKEEKEQSVLKIKEHKAQPTQMPPEPGCILEIVNNVLPQO 89
 1011 NHMDNDDELTPPIYSLKYSDEQ-----LNSGRQSPQNERAARPKHTE--DEIQSEQ 1064
 QY 90 KQNEKEPQTLHQEFYKQVKKMGESQO--QEQKQDAPTCGICHTKTFADCGH 146
 Db 1065 RQSRNQ-----TTPYTESTDDKHLKFPHPGOE-----CVSPYRSKANGSET 1111
 QY 147 NCSTCQTFCAKCGRVSLSRKNVMVNCMLCKQOEILTKSGAMFYNSGNTLQDPQKY 206
 Db 1112 N-----RVGKNHGINQNVQSILC-----QEDDYEDDKPTNYSERYSEEQHEE-- 1154
 QY 207 PGLRNEAPQEKKAKLHEQPOFGAPGDLSPVAVKGRAGHGLTRDITKNSGVKHQIA 266
 Db 1155 -----EERPIYSLIKYDEKRNHVDQPIDYSL-----KTA 1183
 QY 267 SDMSDRKRSFVSVDQNRKYQSEEREDYSQVPSDGTMPSPDYADRRSQREPQFYE 326
 Db 1184 TDIPSQKQSFSSKSSSGQSSKTE-----HMSSENTSTPSSNARKQNLHMSAQ 1236
 QY 327 EPGHLNYSNRGRHSHKE---YIVD-----EDVE 335
 Db 1237 SRSQPOKAACTCKVSSINQETIQTYCVEDTPICFKSCSLSLSSAEDETGNOTQEAD 1296
 QY 356 SRDEXERQREERYQARYRSDPNLARYPKQPYEFQMRTHAEVSARHERHRSDVSLAN 415
 Db 1297 SANLIQLEIKETIKTSRANDEVSEYPAVSQHPRTKSSRLQGS--SLSSSARAKAVEFSS 1355
 QY 416 AELEDSRISLMDRP-----SROSVSERRAMENQSVSMERTREAO---- 459
 Db 1356 GAKSPKSGAQTPKSPREHYVQETPLMFSCRISVS--SLDSFEERSLASSVQSPCSC 1411
 QY 460 -----GOSSTPQRTSNISPPTRPSPIPLDRPMKRRADSLRKHLLDSSAVAKTKRREK 513
 Db 1412 MVSGLISSPDLDPSPGQTMPPSPRSKTPP-----PQTA--QTKRE- 1450
 QY 514 METMLRNDLSLSDQSESVRPPRPKSKKGGKMRQVSLSSSEEL-----ASTPYTSC 568
 Db 1451 ---VPKNAPTAEKRES-----GPKQAQVNAVAQRYQVLPDADTLHLHATESTPDGFCSC 1501
 QY 569 D-----DYEL-----ESESVESEKGSQ--KGRKRT--SEOG 595
 Db 1502 SSSLSALSLDEPFIQKQVDELRTMPYQENDNGNETESEQPKESNENQEKAEKTIIDSEK 1561
 QY 596 VLSDSN-----TRSERQAKRMVYGGHSLSEDLWSEPDITKDSGVDTCCS 639
 Db 1562 LLDDEDDDDIEILLECIIISAMPTKSSRKAKPAQATASKLPPVARKRPSQ-----PVYK 1615
 QY 640 TTLNEHSHSDKHPYMQPSKXGDR--IGRILLN-----KRLDGSVPRDSGAMGL 650
 Db 1616 LIPSONRLQPKH--VSFPGDDMPRYVCEGTPINSTATSLSDLTIESPPNLAAGEV 1674
 QY 691 KVVGGKATSEGLCAFITKVKKSLADYVGHLPDGDEVLEWNGRLIQAGTFEEVYNIILE 750
 Db 1675 R--GG--AQSG-----EFERKDTIPTEG--RSTDEA-----QGCTSSVTITPELD 1713
 QY 751 SKPEQVELV-----SRPIGD-----IPRIPDSTHAOLESSSSSFESOKMDRPSISVT 799
 Db 1714 DNKAEEGILAEICINSAMPKGSKHPFRVYKIKIMDQVQ--QASASSSAPNKNQDLGKKKKPT 1772
 QY 800 SPMSPGMLRDVQVPLSGQLSLKMFQKVGHOLIVTILGAKDLPSREDGRRPNRYVATYFL 859

Db 1773 SPVKP-----IPQNTERTRY-----RKNDSKNNLNAAERVF 1804
 QY 860 PDRSDKRRRTKTVKKTLEPK-----WNQTFISPVH-----RR 893
 Db 1805 SDNDSKKQNLKNNKSKDFNDKLPNNEDRVGRSFAFDSPPHATPIEGTPYCFSRNDSLSSL 1864
 QY 894 EFRERMLEITLMOARRESESEFLGELLILEETALLDDEPHWYKIQHDVSSLPRLP-- 951
 Db 1865 DFDDDDVDLDR--EKAELRKAENKRESEKAYSHLETLSNQSANTQA--TAKQPTNRQ 1921
 QY 952 PPSYLPRLP-----HGESPTRLQr-----SKRISDEVSVDYDCE-- 987
 Db 1922 PKPILQKSTFPQSSKDLIPDGAATDEKIQFATIENTPVCFSNHSLSLSDIDQENNNK 1981
 QY 988 -----DGVGVSDYRHHNG-----RDLOSITLSPEDVQVSSNHC 1020
 Db 1982 ENERIKETEPPDSQGEPSKQASGYAKSFHVEDTPVCFSSNSSLSSLSIDEDLLQEC 2041
 QY 1021 SPSSPHRVD---VIGTRSMSPSAPPQRVNVEQHGRTATGHNTISMD--RHRVMD 1076
 Db 2042 ISSAMPKRRKPSRLKGNEKHSF-----RNM--GGILGEDTLTKDQRPDSEHGLSPD 2094
 QY 1077 HYSDDRDRDCEAAD-----RQ-----PYHRSRSTE 1101
 Db 2095 SENDFKALQEGANSIVSLSLHQAANAACLSRQASDSDSLILSLKSGISLGSFPHITPDE 2154
 QY 1102 QRPL-----ERTTSSRSERPDIN----- 1122
 Db 2155 EKPTSNKGPRLKPGKESLTETKIESEKIGKQKVYKSLITGKVRNSSEISQMKQ 2214
 QY 1123 -LMSMSLMTGRAPSPALSRSHPRGVS-----QTSPSSTPGGRGRQRLPOLP 1174
 Db 2215 PLQANMSISGRPMIHPGRNSSSTSPVSKKOPPLKTPASKSPSEG---QVATTSP 2270
 QY 1175 KGLTERAMDIIEERNOMKILNKYQVAGSDPRLQDYHSKYRSG---MDPHRGAD----- 1226
 Db 2271 RGAKPSYKSELSPARQ-----SQIGSS-----KAPRSGRSDTPSRPAQPLSR 2318
 QY 1227 TVSTKSSDSVDVSAVSRTSSASRFSSTSYMSVQSERPRGNRKISVFTSKMQRQMGVS 1286
 Db 2319 PIQSPGRNRSIPGRNGISLPPKLLQLPRTSSPTASTKSSSGKMS--YTS--PGROM--S 2373
 QY 1287 GKNTKTSISGDMCSLEKNQSGSDTAVGALGTSKKRRBSIGAKMAITGLSKRSRSA 1346
 Db 2374 QQNTLKQGTGLSKNASLIPRSASAKGLNQNNGANGANK-----VELSRMSTFK 2422
 QY 1347 SOLSQTEGGGKTL-----RSTVQRSTETGLAVEMRNMTQASREST----- 1388
 Db 2423 SSGSESDRSERPVLVROSTFIKEAPPTLRKRLLESASFESLSPSSRPASPTRQQAQTPV 2482
 QY 1389 -----DGSNMSYSS 1397
 Db 2483 LSPSLPDMSLSTHSS 2497

RESULT 4
 US-08-289-548A-7
 ; Sequence 7, Application US/08289548A
 ; Patent No. 5648212
 ; GENERAL INFORMATION:
 ; APPLICANT: ALBERTSEN, HANS
 ; APPLICANT: ANAND, RAKESH
 ; APPLICANT: CARLSON, MARY
 ; APPLICANT: GRODEN, JOANNA
 ; APPLICANT: HEDGE, PHILIP J.
 ; APPLICANT: JOSLYN, GEOFF
 ; APPLICANT: KINZIER, KENNETH
 ; APPLICANT: MARKHAM, ALEXANDER F.
 ; APPLICANT: NAKAMURA, YUSUKE
 ; APPLICANT: THLIVERIS, ANDREW
 ; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 ; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,548A
FILING DATE: 12-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-289-548A-7

Query Match 3.1%; Score 258.5; DB 1; Length 2842;
Best Local Similarity 17.9%; Pred. No. 2.7e-10;
Matches 304; Conservative 238; Mismatches 618; Indels 535; Gaps 69;

QY 30 SHLTDEERKIIILAVMDROKKEEKEOSVLIKIEKHKAOPTMPFPFSGITELVNNVLOPQQ 89
DB 1011 NIMDNDGELDPINYSKLSDBEQ----LNSGRSPSQNRMARPKHITF--DEIKSEQ 1064
QY 90 KQPKNEPOTKLHQEFMYKEOVKKMGESQ--QEOBKGAAPTCGICHTKFPADCGH 146
DB 1065 ROSRMS-----TTPYVTESTDDKHLKFQPHFGQE-----CVSPYRSGANGSET 1111
QY 147 NCSYQOTKRCAGCGRVLSRSKKVMVYVLCRKQOELTKSGAMFYNSSGNTIQQPQKV 206
DB 1112 N-----RVGSHNGINQVNSQLC-----QEDDYEDDKPTNSERYSEEQHEE-- 1154
QY 207 PGLRNEAPQEKKAKLHEDQFOGAPGDLSPVAVKRGRAHGLTROPTIKNGSGVKQIA 266
DB 1155 -----EERPTVYSIKYNEKRVHDPIDYSL-----KYA 1183
QY 267 SMPSDRKSPSVSRDQNRREYQSEEREDYSQYVPSDGTMPRSPDYARRSQEPQFYE 326
DB 1184 TDPSSQKQSFSSKSSQSSQSKTE-----HMSSESENTSPSSNAKRONOLHPSSAQ 1236
QY 327 EGHNLVRSNRRGRHSHKE---YIVDD-----EDVE 355
DB 1237 SSSGQPKAKATCKVSSINQETIOTYCEVEDTPICFSSCSLSSLSAEDETGICNOTTDEAD 1296
QY 356 SHDEYERQREEEYQARYSPDLARYPVKPPQYEQMRIHAESRARHRRHSDVLAN 415
DB 1297 SANTQIDAIKIKKIGTRSMADPEYSEVPASQHPRTKSRIGGS--SLSSSEARHKAIVFSS 1355
QY 416 AELSDRSISLLRMDRP-----SRQSVSERRAAMNORSTYSKERTREAO---- 459
DB 1356 GAKSPSKSGAQTTPKSPPEHYVQETPLMFSCRSTSVS-----SLDSFESKSIASSVQSEPCSG 1411

QY 460 -----GQSYQARTSNHSPPTPRRSPIDLRPDMRRADSLRKHHLDPSSAVRKTREK 513
DB 1412 MVSGLISSDLPDSEFGQTMPPSRKSTPPP-----PQOTA--QTKRE-- 1450
QY 514 METMLRNDLSLSDQSESVRPPPRPHKSKGKGKQVSLSSSEEL-----ASTPEYTCG 568
DB 1451 ---VPKNKAPPAEKRES-----GPKQAAVNAVQVGLPADVLLHPTATESTPDGFGC 1501
QY 569 D-----DVEL-----ESESVSQKGSQ--GKKRKT---SEGG 595
DB 1502 SSSLSALSLEPFIQKDELIRIMPVQENDNGNETESQPKSNENQKEAEETIDSEKD 1561
QY 596 VLSDSN-----TRSERQKKRMYYGCHSLEEDLEWSEPOIKDSQVDTCSG 639
DB 1562 LDDSDDDDDIIEECIISAMPTKSSRKAKRAQASKLPPPVARRPSQL-----PVK 1615
QY 640 TTLNEHSHSKHPVYTWOPSKDGRLL--IGRIILN-----KRLKGSVPKDSGAMLCI 690
DB 1616 LLPSONRLQPKH--VSFTPPGDDMPRVYCEGTPINFTSTSLSDLTIESPPMELAGEGV 1674
QY 691 KVGSGKMTESGRLCAFITKVKKGSILADIVYGLRPDDEVLEMNGLLQATFEFYNIILE 750
DB 1675 R--GG--AQSG-----EFEKRTIPIEG--RSTDEA-----QGGKTSVTIPELD 1713
QY 751 SKPEQVELVY-----SRPIGD-----IPRIDSTHAOLESSSSSFESQKMDRPSISVT 799
DB 1714 DNKAEEGDIILAEICISAMPKKGSHKPRFKKIMDOVQ--QASASSAPKNNQDGLKKKKFT 1772
QY 800 SPMSPGMLRDVPOFLSGQLSIKLMFDKVGHQLIVITLAKDLPSREDGPRNRYKIYEL 859
DB 1773 SPVKP-----IPQNEYETRV-----RKNADSKNNLNAERYF 1804
QY 860 PDRSDKNKRTKTYVKTLEPK-----NQPTIYSPVH-----RR 893
DB 1805 SDNKSQKQNLKNNKSKDNDKLPNNEDRVYSGFADSPHHYPIEGTPYCFSRNDSLSTI 1864
QY 894 EFERMLETTLMDQARVRESESEFIGEILIELETALLDDEPMYKQGTQDVSLPLPR-- 951
DB 1865 DFDDDDVDLISR--EKALRKAKENKSEAKVTSTELTSMQASANKQA--IAKQPINRQO 1921
QY 952 PSPYLPRROL-----HGSEPTRLQR-----SKRISDSEVSDYDCE--- 987
DB 1922 PKPILQKQSTPQSKDLPDGAATDEKLQNAIENTPYCFSHNSLSLSDIDQENNNK 1981
QY 988 -----DGVGVSDYRNG-----RQIQSSTLSVPDQVYSSNHC 1020
DB 1982 ENERIKETEPDQSPQSPQASGYAPKSFHVEDTPVCFSRNSSLSTLSDSEDDLQEC 2041
QY 1021 SPQSGPHRYVD---VIGRTFRSPSPAPRQORVNEQGHGRTRAGHYNTISRMQ--RHRVMD 1076
DB 2042 ISSAMPKRRKRSRLGDNKEKHSF-----RNM--GGILGEDLILDLKDQRPDSHGLSPD 2094
QY 1077 HYSSDRDRDCEAAD-----RO-----PYHRSRSTE 1101
DB 2095 SENFDWKAIOEGANSIVSLHQAAAAACLSROASSDSDSLILSLKSGISLQSPHLTPDQE 2154
QY 1102 QRPLL-----ERTTSSRSSEPDN----- 1122
DB 2155 EKPTFSNKGPRILKPGKESTLETKKIEESKGIKGGKYYKSLITGKVSNSIISGOMQ 2214
QY 1123 -LMRSPSLMTGRSAPSPALSRSHPRTGSV-----QTSPTSGTQARGQQLPOLP 1174
DB 2215 PLQANMPSTISRGRTMIHIFGVANNSSSTSPVSKKGPPLKTPAKSKSSEG---QTATISP 2270
QY 1175 KQTLERSAMDIERNRQMLINKYKOVAGSDPRLQEDYHSHKYSRG---WDPHRCAD----- 1226
DB 2271 RGAKPSVKSSELSPARQT-----SQIGSS-----KAPSRSGSDSPSPRAQOPLSR 2318
QY 1227 TVSTKSSDSQSDVSAVSTSSASRPSSTSYMSVQSERGRNKKIIVTFYSKMNQRMQVS 1286
DB 2319 PLOSPGRNISPGRNGISPPNKIJSOLPRTISSPSTASTKSSGSGKMS--YYS--PGRQM--S 2373


```

OY 569 D-----DVEL-----ESESVEKGDQ-KGRKRT--SEQG 595
Db 1503 SSSLALSLDEPFIQKVDLRIKMPVOENDNGMETSECEKESNENDEKAEKTIIDKED 1562
OY 596 VJSDSN-----TREROKKRMYYGGHSLBEDLEWSEPPQIKDSVDTCCS 639
Db 1563 LDDSDDDDEIILEECIIISAMPKSSKAKKPAQATASKLPVPARKBSQI-----PVYK 1616
OY 640 TILNEHSHSDKHPVTWOPSKDRL--IGRIILN-----KRLKGSVPBDSGAMJL 690
Db 1617 LIPSONRLOPQKH-VSFTEGDDMPRYCYEBGTINFSTATSLDLETSPNELAAGECV 1675
OY 691 KYVGGKMTESGRICAFITKVKKSLADYVGHLPBGDEVLEWNGRLLOGATFEVYNTILE 750
Db 1676 R--GG--AOSG-----EEREKDTIPTEG--RSTDEA-----OGKRTSVTIELD 1774
OY 751 SKPEQVELVY-----SRPLGD-----IPRIDSTHAOLESSSSSEOSKMRPSISVT 799
Db 1715 DNKAEGBDILAEICINSAMPKSGKSHKPPRYKKIMDOVO-QASASSAPKNQOLDGKKRPT 1773
OY 800 SPMSGMLDVPQFISGOLSIKLPFDKVGHLVTLIGAKDLPSREDGRPNPVKITYFL 859
Db 1774 SVYKP-----IPONTYTRV-----RKNADESKNMLNAERV 1805
OY 860 PDRSDKNKRRITVKKTLBPK-----WNQTFYSPVH-----RR 893
Db 1806 SUNKDSKKONLNKNSKDFDKLPNNEDRVKGFAPDSPHHYTPIEGTPYCFGRNDSLSL 1865
OY 894 EFRERMLETITLDQAVRESEFEIILEETALIDDEPHWKIOTHDVSSLPDR-- 951
Db 1866 DDDDDVDVLSR-EKALIRKAKENKESAKVTSHTELSNQOSANKTQA--IAKPIRNGQ 1922
OY 952 PBPYLPBRQL-----HGESPIRRLQ-----SKRISSEVSDVCE---- 987
Db 1923 KPPILOKOSTPFOSSKDIPTDKGATDEKLONFALIENTPVCFSHNSSLSLSDIDQENNK 1982
OY 988 -----DGVVVSRYRHNG-----RDLOSSTLVSPEQWSSNHC 1020
Db 1983 ENEPKETEPPSOGSPKPOAGVAPKSHVEDTYPCEFRNSSLSLSDSEDDILOEC 2042
OY 1021 SPGSGPHRYD---VIGRTMSWSPSAPPORNVQOHRGTRATGHYNTISMD--RHRVMD 1076
Db 2043 ISSAMPKKKRPSRLKGDNEKHP-----RNM-CGILGEDLTLDKIDQRPDSHSGSPD 2095
OY 1077 HSSSDRRCLEAD-----RQ-----PYRKSSTE 1101
Db 2096 SENFPMKALIOEGANSLVSSLIHQAAAAACISROASSDSLSLSKSGISLGSPEHLTPDQE 2155
OY 1102 QRPPL-----ERTTTRSRSSERPDTN----- 1122
Db 2156 EKPLTSSNKGPRILKPEKSTLEKKTIESEKGIKGGKVKYKSLITGKVRNSLSGOMQ 2215
OY 1123 -LMRSMPSLMTGRAPSPALSRSHPTGVS-----OTSPSSTPGTGRROLPLP 1174
Db 2216 PLQANMPSTISRGTMHIGVNRSSSTSPVSKKGPPLKTPAKSPSEG-----QTATTP 2271
OY 1175 KGTLEASAMDIERNOMKLNKYKOVAGSDPRLEQDYHKSVRSG--WDHHRGAD----- 1226
Db 2272 KAKKPSVKSELSEVAVOT-----SQTIGSS-----KAPSRSGSRDSTSPRPQOPLSR 2319
OY 1227 TVSTKSDSDVSVAVSRVTSASRFSSTSYMSVQSERPPGNRKISVFTSKMOROMGV 1286
Db 2320 PLOSPORNISIPGRNISPKNLSQLPRTISSPSTASTKSSGSKMS-YTS--PGROM--S 2374
OY 1287 GNLTKSTISIGDMCSLEKNDGQSDTAVGALCTSGKKRRSSIGAKVAIVGJSRKRSRA 1346
Db 2375 QONLTKOTGLSKNASSIPRSESASKGLNOMNNGANK-----VELSRMSTK 2423
OY 1347 SOLSOTEEGGGKKL-----RSTVORSTETGLAVEMKMMMTQOASRESI----- 1388
Db 2424 SSGSESDRSERPVLVROSFTIFKEAPSPTLRKLEASAFESLSPPSRPASPTRSOAOTPV 2483
OY 1389 -----DGSMSYSS 1397

```

```

Db 2484 LSPSLPDMSLSTHSS 2498
RESULT 7
US-08-452-655B-7
; Sequence 7, Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,655B
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; US-08-452-655B-7

Query Match 3.1%, Score 258.5; DB 1; Length 2843;
Best Local Similarity 17.9%; Pred. No. 2.7e-10;
Matches 304; Conservative 238; Mismatches 618; Indels 535; Gaps 69;

OY 30 SHLTFEERKIILAVMDROKKEEKEQSVLKIKEEKAQPTOMFPSPSGITELVNNVLOPOQ 89
Db 1012 NHMDNDDELPTPIVSLKYSDEQ---LNSGRQSPSONERARKKHILE--DEIKOSIQ 1065
OY 90 KQPNKEPOTRLHOQFEMYKKEQVKKMGESQO---QOEKGDAPTGICIKTKTFADGCGH 146

```

```

Db 1066 ROSRMO-----TTYTESTDDKHLKFPHFQOE-----CVSPYRSKANGSET 1112
QY 147 NCSYQOTFCARCGRVSLRSKVMVYCNLCRKQOEILTRSGAMFNSSNTLQOQDOKY 206
Db 1113 N-----RVGNSHGINOVNOSLQ-----QEDDYEDDKPTNYSERYSEEOHHE-- 1155
QY 207 PGLRNEBAPOEKKAHLHEOPQOGAPGLSVPAVEKGRAGHLTRDQTLKNGSGVHOLA 266
Db 1156 -----EERPTVYSIKYNEKRVHVDOPIDYSL-----KYA 1184
QY 267 SDMPDRKRSPSVSDQNRREYQSEEREDYSQYVPSDGTMPNSPSDYADRRSQREPOFYE 336
Db 1185 TDIPSOKOSFESFKSSGOSKTE-----HMSSENTSTPSSNAKRONOLHPSAQ 1237
QY 327 EPGHLYNDSNRRGRHRSKE-----YIVD-----EDVE 355
Db 1238 SRSGOPQARATCVSSINQETIOTYCEVEDTPICFSCSSLSLSADEIGCNOTTOEAD 1297
QY 356 SRDEXERQREBEYQARYSDENLARYPVKPOPYEOMRTHAEVSRARHRRSDVSLAN 415
Db 1298 SANTIQIAEIKETIGTRSNEDVSEVPASQHPRTKSSRLQGS--SLSSSARKKAVEFSS 1356
QY 416 AELEDSRISLMDRP-----SRQSVSERRAMENORSYSMERTREAO-----459
Db 1357 GAKSPKSGAQTPKSPREHYOETPLMFSCRYSVS-----SLDSFESRIASSVOSEPCSG 1412
QY 460 -----GGSYTPQRTSNISPPTRPRSPITPLDRPMRADSLRKOHLLDSSAVRKTREK 513
Db 1413 MWSGIISPSDLPDSFGQTMPPSRKTPP-----PQTA--QTKRE- 1451
QY 514 METMRNLSLSDQSESVRPPPRPKSKGKMGROVSLSSSEEL-----ASTPEYTC 568
Db 1452 ---VPKNAPTAERES-----GPKQAANVAQVQVYLPADOTLLHATETSTPPGFCSC 1502
QY 569 D-----DVEL-----ESESVESEKGSQ-KGKRT--SEOG 595
Db 1503 SSSLALSLEBFIQDVELRIMPVQENDNGNETESQPKESNENQEKAEKXTIDSEKD 1562
QY 596 VLSQDN-----TRSEROKKRMYYGSHLEEDLEMESEPOIKDSGVTCSS 639
Db 1563 LLDSDDDDIETLEECIISAMPTKSSRAKKRPAQTASKLPPVARRKPSUL-----PVYK 1616
QY 640 TTLNEHSHSDKHPYTWQPSKDGRL--IGRILN-----KRLKDGSVPRDSGAMGL 690
Db 1617 LLPSGNRQLOPKH-VSTFRGDDMPRYVCYEGTRPINFSTLSDLTIESPPLAAGEVY 1675
QY 691 KVGSGKMTESGRCAFITKVKKGLADTVGHLRPGDEVLEWNGRLQATFEVYNIILE 750
Db 1676 R--GG--AOSG-----EFEKRODTIPTEG--RSTDEA-----QGGKTSVTLPELD 1714
QY 751 SKPEQVELV-----SRPID-----IPRIPISTHAOLESSSSSFESOKMDRPSISVT 799
Db 1715 DNKAEGGILAEICINSAMPKSGSHKPRFYKKTMDQVQ--QASASSSPANKNOLDGKKKPP 1773
QY 800 SPMSPGMLRDVPOFSGOLSTIKLWFDKVGHOILVTLGAKDLPSRECDGRPNRYVAYILE 859
Db 1774 SPVAP-----IPONTETRYV-----RKNAOSKNNLNAERVF 1805
QY 860 PDRSDKNRRTKYVKLTLEPK-----WNQPIYSPVH-----RR 893
Db 1806 SDNKSOKKONLKNNKSKDNKLPNNEDVRGSEFADSPHHYPIETCTPYCFSSNDLSL 1865
QY 894 EFERMELEITLMDOKARVEESEFLEIILETALLDDEPMYKIQTHIVSLEPLR-- 951
Db 1866 DFDDDDVLSR-EKALRKAKENKSEAKVTSHTELTLSNOQSANQQA--IAKQPIINRQO 1922
QY 952 PSPYLPRLQ-----IGESPTRLOR-----SKRISDSVSDVDC-----987
Db 1923 PKPILOKSTPQOSKQDLPDKGAATDEKIQNAIENTPVCFSHNSLSLSDIDQENNNK 1982
QY 988 -----DGVGVVSDYRIHG-----RDLQSTLSVPEQVMSNHC 1020
Db 1983 ENPIKETEPDPSQGEPSKPAQASGFAPKSFHYEDTIPVCSFRNSLSLSDSDLDLQGC 2042

```

```

QY 1021 SPGSGPHRVD---VIGTRMSPSAPPPQNRVNEGHRGTRATGHYNTISRMD-RHRVMD 1076
Db 2043 ISSAMPKKKRSRLKGNESHSP-----RNM-GGILIEDTLDKDIQRPSEHGLSPD 2095
QY 1077 HYSRDDRDCEAD-----RQ-----PYHRSRSTP 1101
Db 2096 SENFDKMAIQGANSIVSLLQAAAAACLSDQASSDSISLSKGSISLGSFHLTPDOE 2155
QY 1102 QRPIL-----EKTTRRSRSERPDTN-----1122
Db 2156 EKPPTSMKGPRLKPGEKSTLETETKIESESGIKGKKVYKSLTGKYSNSEISGOMQ 2215
QY 1123 -LMSRMSLMTGRSAPRPPALSRHPTGVS-----QTSSTSTPGGRRQRLPOLP 1174
Db 2216 PLOANMPSISGRTHIHPGVNRSSSTSPVSKKGPPLKTPAKSPSEB--QATATSP 2271
QY 1175 KGTLERSAMDIEERNQMKLKNKYOVAGSDPRLDQDYHSKYRSG--WDPHRGAD-----1226
Db 2272 RGAKPYSKSELSPVAKQ-----SOIGSS-----KAPSGSGRSDTPSKPDAQPLSR 2319
QY 1227 TVSTKSSDSDVSDYSAVSRTSSASRFSSTSYNSVOSEPRGRNKISVFTSKMQRMGVS 1286
Db 2320 PLOSPGRNISIPGRNGLISPPMKLSQLPRTSPSTASTKSSGSGKMS-YTS--PGROM--S 2374
QY 1287 GKNTKTSTISGDMCSLEKNGSOSDTAVGALGTSKGRRRSIGAKKMAIVLSKRSRA 1346
Db 2375 QONLTQOTGLSKNASIIPRSESASGKLNOMNNGANK-----VELSRMSTK 2423
QY 1347 SOLQOTEGCGKKL-----RSTVORSTETGLAVEMNMWTROASRST-----1388
Db 2424 SSGSESDRSEPRVLVROSTFTKEAPSPILRKLREBSASFESLSPSRAPSPRQAQTPV 2483
QY 1389 -----DGSNMSYSS 1397
Db 2484 LSPSLPDMSLSTHSS 2498

```

```

RESULT 8
US-08-450-582-2
; Sequence 2, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THILVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Wilcoff, Ltd.
; STREET: 1001 G Street, NW
; City: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,582
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```



```

Sequence 7, Application US/08450582
Patent No. 6114124
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THILVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESS: Banner & Wilcoff, Ltd.
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450,582
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/452,655
FILING DATE: 25-MAY-1995
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9299
TELEFAX: 202-508-9100
INFORMATION FOR SEO ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
US-08-450-582-7
Query Match 3.1%; Score 258.5; DB 3; Length 2843;
Best Local Similarity 17.9%; Pred. No. 2.7e-10;
Matches 304; Conservative 236; Mismatches 618; Indels 535; Gaps 69;

```

```

Db 1113 N-----RVGNHNGINQVNSQLC-----QEDDYEDDKPTNYSERYSEEQHEE-- 1155
QY 207 PGLRNEEAPQEKAKLHEPOFOGABGDLSPAVEGKRAHGLTRQDTIKNGSVKQIOA 266
Db 1156 -----EERTNYSIYNEKRRHVDPIDYSL-----KYA 1184
QY 267 SDMPSDKRKRSVSRQONRYEEOSEEREDYQVYPSDGTMPRSPDYADRROREPOFEY 326
Db 1185 TDIPISSOKOSFSPSKSSGSSKTE-----HMSSESTSTYSNAKRONQLHPSAQ 1237
QY 327 EPGHLNRYDSNRGRHRSKE-----YIVDD-----EDVE 355
Db 1238 SRSGQPOKATCKVSSINGDTIOTYCEVEDPICHSCSSLSLSSADEIGCNGTQOAD 1297
QY 356 SRDEYRQREERYQARYSDPNLARYPKVPEYEDQRIHAEDVSRARHERHSDVSLAN 415
Db 1298 SANTLQIAIEIKKIGTRSAEDPVEYPAVQHPPTKSSRLQGS-SLSESARHRAVERSS 1356
QY 416 AELEDSRISLLRMDR-----SRORSYSEBRAAMENORSYMERETREAO----- 459
Db 1357 GAKSPKSGAQTQPKSPPEHVVQETPLMFSCTSY-----SLDSESSSIASSVSEPCSG 1412
QY 460 -----GQSSYPORTSNHSPPTERRSPPLDRPDMRRADSLRKQHNLDPSAVKRTREK 513
Db 1413 MVSGLTSPDLPSPQGTMPSPKSKTPEP-----PQTA-QTKRE- 1451
QY 514 METMLRNDLSLSDQSVPRPPRPHKSKGKMYOVSLSSEEL-----ASTPEYTC 568
Db 1452 ---VPRNKAPTAERES-----GPKQAAVNAVQRYVLPDADTLHFAETSPDDESC 1502
QY 569 D-----DVEL-----ESESVEKGDQO-KGRKT--SEOG 595
Db 1503 SSSLALSLEPFIQKDVLEIRIMPVQENDNGNETSEQKESNENQEKAKTIDSEKD 1562
QY 596 VLSDSN-----TRSERQKKMYGGHSLBEDLEWSEPOTKDSQVDTCCSS 639
Db 1563 LDDSDDDDEIELECIISAMPTKSSRAKPAQTAKLPPVARKPSQL-----PYK 1616
QY 640 TTIINEHSHDKHPYVQSPKDGDL--IGRLLN-----KRLKGSVPRDGAHGL 690
Db 1617 LLPSONRLOPKH-VSEPTDDMPRVYCVGCTPINFSTATSLDITLIESPNEIAGEV 1675
QY 691 KVVGKMTESGRICAFITKKGSLADTVGHLRPGDVLNNGLLGATGEYYNIILE 750
Db 1676 R--GG--AOSG-----EKKRDTPTG--RSTDA-----OGKTSSTIPELD 1714
QY 751 SKPEQVELVY-----SRPGD-----IPRIDPSTAOLESSSSPESQMDPISVT 799
Db 1715 DNKAEGDILAEICINSAMPKCKHKKPRVKKINDQV-QASASSASANKKQDCKKKKPT 1773
QY 800 SPSPGMLRDVPOFLSGQLSIKLMFDKVGHLVTLIGANDLPREDGPRPNPYVKIYFL 859
Db 1774 SPVKP-----IPQNTERTRYV-----RKNAWSKNMLNAERVF 1805
QY 860 PORSCKNKRRTKVTKLTLEK-----WNQFTYSPVH-----RR 893
Db 1806 SDMKDSKKOMIKNNNSKDFNDKLPNNEDRVGRSFAFDSPHHTPTTEGPPYCFSRNDSLSTL 1865
QY 894 EFERMLEITLMDQARRESESEFLGELLEETALDDEPWHYKLTQHDVSLPLR- 951
Db 1866 DFDDDDVDLSR-EKAEIRKAKENKESAKYTSHTLSTNOOSAKTYA--IAKPIPNRGQ 1922
QY 952 PSEPLPRQL-----HGSPTRRIOR-----SKRISDEVSVDYCE--- 987
Db 1923 PKPTIQKSTFFPOSSKDIPDRCATATEKLONFAIENTPVCFSHNSLSLSDIDQENNNK 1982
QY 988 -----DGVGVYSDVRHNG-----FDLQSSVLSVDEQVMSNNHC 1020
Db 1983 ENEPITEPEPPDSQGESEKQOASGVAPKSFHVEDTPVCFSRNSSLSTIDSEDLLOEC 2042
QY 1021 SPFGSPHRYVD--VIGRTRSMSPSAPPPOPNVNVQGHGATGATGHYNTISRM-D-HRYVMD 1076
Db 2043 ISSAMPKKKKPRSLKGDNEKHP-----RNM-GGILGEDTLTDKTIQRPDSHGSLSPD 2095

```

```

OY 1077 HXSDRDRCEAD-----RQ-----PYHRSKTE 1101
Db 2096 SNFPMKAIQEGANSTVSSLHQAACLSRQASDSDSLSLKSGISLSPHILTDQE 2155
OY 1102 QRPRL-----ERTTGRSSSERPDTN-----1122
Db 2156 EKPFTSNKGRILLKPGKSTLFTKRIEESKGIKGGKVKYLITGKRSNSETSGOMQ 2215
OY 1123 -LMRSMPSLMTGRSAPSPALSHPRTGVS-----QSPSPSTPGTGKRGQLPQLP 1174
Db 2216 PLOANNPSTISRGTMHIGVNRSSSTSPVSKKGPPLKTPAKSPSEG---QTATPSP 2271
OY 1175 KGLTERSAADIEERNOMKLNKYQVAGSDPRLEODYHSKYRSG---WDPHAGAD-----1226
Db 2272 RBAKPSVKSELSPVARTQ-----SQIGGSS-----KAPSRGSRNSTSRPAQQLSR 2319
OY 1227 TVSTSSSDSDVSDVAVSTSSASRFSSTYSMSVOSERPGRNKRISVFTSKMONRGVVS 1286
Db 2320 PLOSFGNRNISFGNRNISPPNKLSQLPRTSSPSTASTKSSSGSKMS-YTS--PGRQM--S 2374
OY 1287 GKNLTKSTISIDMCSLEKNDGSDPTAVGALGTSGKKRRSSIGAKMVAIVGLSKRSRA 1346
Db 2375 QONLTKOTELSKNASSIPRSESASKGLMOMNNGANKK-----VELSRMSTK 2423
OY 1347 SOLSOTEGGKKL-----RSIVQSTETGTGLAVERMMMTROASREST-----1388
Db 2424 SSGSDEDRERVLVQSTFIKEAPSPILRKLEBSASFSLSPSRPASPTPSQAQTPV 2483
OY 1389 -----DGSMSYSS 1397
Db 2484 LSPSLPDMSLSTHSS 2498

```

```

RESULT 10
US-08-821-355A-7
; Sequence 7, Application US/08821355A
; Patent No. 5851775
; GENERAL INFORMATION:
; APPLICANT: Barker, Nick
; APPLICANT: Clevers, Hans
; APPLICANT: Korinek, Vladimir
; APPLICANT: Morin, Patrice
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; APPLICANT: Sparks, Andrew
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC
; TITLE OF INVENTIONS: Interact to Prevent Cancer
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,355A
; FILING DATE: 20-MAR-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A
; REGISTRATION NUMBER: 32,145
; REFERENCE/DOCKET NUMBER: 1107.05064
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: 97430 BMB-UT
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2973 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: NO. 5851775e
; US-08-821-355A-7

```

```

Query Match 3.1%, Score 258.5; DB 2; Length 2973;
Best Local Similarity 17.9%; Pred. No. 2.9e-10;
Matches 304; Conservative 238; Mismatches 618; Indels 535; Gaps 69;

```

```

OY 30 SHLFEERKILLAVMDROKKEEKESQVYLIKIEHKAPOTQMPFSGITELVNNVLOPOQ 89
Db 1012 NHMDNDDELDTPIVNSLYKSDQE---LNSGRQSPSONERWARKHITE--DEIKOSBO 1065
OY 90 KQPNKEPOTLIHQEFEMYKQVKKMGESQO---QOEOKGAPPGICHTKTPADGCGH 146
Db 1066 RQSRNQS-----TTYPVYTESTDDKHLKFPNFGQOE-----CVSPYRSRANGSFT 1112
OY 147 NCYVCOTKFCACGCGRVALRSNKVMWVCNLCKRQOEILTKSGAMPYNSGSSNTLQOPDOKV 206
Db 1113 N-----RVGSNHCINQVNSQSLC-----QDDYEDDKPTVYSEYSEEDHEF-- 1155
OY 207 PRGLRNEEAPQEKAKLHEDOPQFQAPGDLVPAVEKGRANGLTRQDTIKNGSVKHQIA 266
Db 1156 -----EERPTNYSIKYNEEKRNVDQPIDYSL-----KVA 1184
OY 267 SDMPSDKRKRSVSDQNRKRYEQSEFEDYSQVPSDGTMPSPSDYADRKQREPOQTE 326
Db 1185 TDIPSSQKQSFSEKSSSGQSGKTE-----HMSSESTNTSPSSNAKRCQMLHPSSAQ 1237
OY 327 EPGHLNVRDSNRRGHRHSEK---YIVD-----EDVE 355
Db 1238 SRSGQPOKATCKVSSINQETIOTYCVEDTPICFGRCSLSSLSAEDIGCNQOTTQED 1297
OY 356 SRDEYERQREERYQARYSDPNLARYPVKQPYEQMRINAEVSRARRHRSDVSLAN 415
Db 1298 SANLQLAIKEIKETIGTRAEADVSEVPVAVSQHPRTKSSRLQGS--SLSSSAHKAVERSS 1356
OY 416 AELEDRIISLRMDRP-----SRORSVSRRAAMENQSYMERIREMO-----459
Db 1357 GAKSPKSGAQTPKSPREHYVQETPLMFSRCTSVS---SLDSFERSIASVQSEPCSG 1412
OY 460 -----GQSSYPORTSNRPTRRSPIPLDRPDMRADSLRKQHHLDPSSAVRKTRER 513
Db 1413 MVSGIISPSDLRPSBQGMTPSRKSTPP-----PPQTA--QTKRE- 1451
OY 514 METMLRNDLSLSDQSESVRPPPRPKSKGKMKQVSLSSSEEL-----ASTPEYTSNC 568
Db 1452 ---VPKNKAPYAEKRES-----GPKQAAVNAVQVQVLPDADVLHFAFSTPDGFSGC 1502
OY 569 D-----DYEL-----ESSVSXKGSQ--KGAKKT--SEOG 595
Db 1503 SSSLALSLDEPFIQKDELIMRPVQENDNGENETESQPKESMNOEAEKRTIDSEKD 1562
OY 596 VLSDSN-----TRSERQKRMVYGGHSLIEDLEWSPQIKNSGYDTCSS 639
Db 1563 LLDDSDDDDIILEECIISAMPTKSSRAKKAQTAASKLPPVAKKPSOL-----PYK 1616
OY 640 TTLNEEHSDBKHVYTWQPSKDGRL--IGRIILN-----KRLKDSVPRDSGAMGL 690
Db 1617 LLPSONRLQPKKH-VSTFRGDDMPRVYVEGPIPIFSTATLSLDTLTPSPRELAGEVY 1675
OY 691 KVGCGKMTESGRICAFITKVKKGLADTVGHLRPGDEVLEWNGRLLQAGATFEVYNIILE 750
Db 1676 R--GG--AQSG-----EFERKDTIPIEG--RSTDEA-----QGGKTSVITPIELD 1714

```

```

QY 751 SKPEQVELV-----SRPIGD-----IPRIPDSTHAESSSSSEFESQKMDRPSISVT 799
D 1715 DNKAEEGIIAECTISAMPKKGSHKFRKKTIMDOVO-QASASSAPNNKNOIDGKKKKT 1773
QY 800 SPMSPGMLRDVPOFLSGQLSIKLMEDKVGHOILVITLAKDLPREDGRPRNPYKIVFL 859
D 1774 SPVKP-----IPONTNEYRTRV-----RKNADSKNNLNAERVF 1805
QY 860 PDRSDKNNRRTKYAKTLEPK-----WNQFTIYSPVH-----RR 893
D 1806 SDNKSQKQNLKNNKSKDNKLPNNEDRVGSAFADSPHHYTPRECTPYCFGRNDLSL 1865
QY 894 EFERMLETILMDQARVREESEFLEIIEFETALLDEPHMYKLOTADVSSLPLPR-- 951
D 1866 DFDDDDVULSR-EKAELEKAKENKSEAKVTSHTELTSTQOSANKTQA--IAKQPINRQ 1922
QY 952 PSPYLPRLRQ-----HGESPTRLQR-----SKRISDESVSDYDCE-- 987
D 1923 PKPILQKOSTFPOSSKDIIPDRGAATDEKLQNFALIENTPVCFSHNSLSLSLSDIDQENNK 1982
QY 988 -----DGAVVSDYRNG-----RDLQSTLSVPEQVMSNNC 1020
D 1983 ENERIKETEPPOSEPKPOASGAPKSFHVEDPVPFCFSRNSLSLSISDEDDLQRC 2042
QY 1021 SPSGSPHRYD--VIGTRRMSGPSAPPPQARVEQGRGTRATGHYNTISRM-D-RHRYMD 1076
D 2043 ISSAMPKAKKKSRLKGDNEKHS-----RNM-GGILGEDLILDKIDRPPDSEHSLSD 2095
QY 1077 HYSDDRDRDCEAD-----RQ-----PYHRSRSTE 1101
D 2096 SENFWMKAIQEGANSIVSLHQAAAAACLSQASSDSISLSLKSGLISGPHLPPDQE 2155
QY 1102 QRPPL-----ERTTRSSSERPDYN-----1122
D 2156 EKPFTSNKGPRLKPKGKSTLETKKIESEKGIKGGKVKYSLITGKVSNSEISGOMRO 2215
QY 1123 -LMRSMPLMTGRSAPSPALSRSHPRTGSV-----OTSPTSGTGRGRLQLP 1174
D 2216 PLOAMPISIKRGTIHTIPGVANSSSTSPVSKKGPPLTPAKSPSECG--OTATTSP 2271
QY 1175 KGLTERSAMDIERNRQKMLNKYQOVAGSDPRLBEDYHSKYRSG--WDPHRCAD----- 1226
D 2272 RGAKPSVKSLSVPAQR-----SQIGSS-----KAPSRSGSDRPSRAQPLSR 2319
QY 1227 TVSTRSSSDVSDVAVRTSASRPSSTSYSVOSERPGRNRKISVFTSKMONROMGVS 1286
D 2320 PIOSGRNSISPGRGISPPNKLSQLPRTSSPSTASTKSSGSKMS-YTS--PGROM--S 2374
QY 1287 GKNLKSTISIGDMCSLEKNDGOSQDTAVGALGTSGKRRRSIGAMVAIVGLSRKRSRA 1346
D 2375 QONLTKQTLGSLKNASSIPRSSESAGLONMNGANKR-----VELSRMSSTK 2423
QY 1347 SOLSOTEGGKKL-----RSTVQRSTETGLAVEMRMNMTROASREST----- 1388
D 2424 SSGSSDSRSEVRLVLRQSTFIKEAPSPILRRKLEESAFESLSPSRAPSPRTSQOQTPV 2483
QY 1389 -----DGSMSYSS 1397
D 2484 LSPSLPDMSLSTHSS 2498

```

```

RESULT 11
US-09-003-687A-7
Sequence 7, Application US/09003687A
Patent No. 5998600

```

```

GENERAL INFORMATION:
APPLICANT: Barker, Nick
APPLICANT: Clevers, Hans
APPLICANT: Korinek, Vladimir
APPLICANT: Morin, Patrice
APPLICANT: Kinzler, Kenneth
APPLICANT: Vogelstein, Bert
APPLICANT: Sparks, Andrew

```

```

? TITLE OF INVENTION: Beta Catenin, TCF-4, and APC
? NUMBER OF INVENTION: Interact to Prevent Cancer
? NUMBER OF SEQUENCES: 11
? CORRESPONDENCE ADDRESS:
? ADDRESS: Banner & Witcoff, Ltd.
? STREET: 1001 G Street, N.W.
? CITY: Washington
? STATE: DC
? COUNTRY: USA
? ZIP: 20001
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FASTSEQ for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/003,687A
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/821,355
? FILING DATE: 20-MAR-1997
? ATTORNEY/AGENT INFORMATION:
? NAME: Kagan, Sarah A
? REGISTRATION NUMBER: 32,145
? REFERENCE/DOCKET NUMBER: 1107.05064
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-508-9100
? TELEFAX: 202-508-9299
? TELEX: 97430 BMB UR
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2973 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: No. 5998600e
US-09-003-687A-7

```

```

Query Match 3.1%; score 258.5; DB 2; Length 2973;
Best Local Similarity 17.9%; Pred. No. 2,9e-10;
Matches 304; Conservative 239; Mismatches 618; Indels 535; Gaps 69;

QY 30 SHLTREERKIIILAVMDRQKEEKEQSVLYKTEHKKAQPTQWPPSGITELVNNVLPQOQ 89
D 1012 NHMDNDGELDTPINYSILKYSDEQ-----LNSGRQSPSONERWARPKHIE--DETKQSEQ 1065
QY 90 KQPEKEPQTKLHQEFMYKQYVKKMGESQO--QOEQKGDAPTCGICHTKTKFADGCGH 146
D 1066 ROSRNQO-----TTYVYTESTDCKLHKFPHFQOQ-----CVSPYRSRGANGSET 1112
QY 147 NCAYCQTKFCARCGRVSLRSNKMVMVNCILRKQOELITKSGAMEFYNSGNTLQOPDOKV 206
D 1113 N-----RVGSNNGINQNVQSILC-----QEDVEDKPKTNYISRYSEEBQHEE-- 1155
QY 207 PGLRLNEAPQEKAKLHEDQFOGAPGDSVPAVEKGRAGHLTRQDTIKSGYKHOIA 266
D 1156 -----EREPYNYSIKYNEEKRHVDQPIDYSL-----KYA 1184
QY 267 SDMPDRKRSPSVSRDQNRREYQOSEEREDYSQYVPSDGTMPRSPDYADRRSOREPQYE 326
D 1185 TDIPISSQKOSTFSTKSSGOSKTE-----HMSSESTSTPSNAKKRONQLHPSAQ 1237
QY 327 EPGHLNRYDSNRGRHRSKE-----YIVDD-----EDVE 355
D 1238 SRSGQPKAATCKVYSINQETIQTYCVEDPTICFSRCSLSLSLSAEDEIGCNOQTQAD 1297
QY 356 SRDYEERQREETQOARYRSPNLAIRYPVKQPYEOMRTIAEVSRARHERHSDVSLAN 415
D 1298 SANTLQIAETKEIKGTGSAEDPVSEVPAVSOHPRTKSSRLQGS--SLSESAKRAKVEPSS 1356
QY 416 AELEDSITSLILMRDP-----SRQRSVSERRAAMENGRSYSMETREAO----- 459

```

```

Db 1357 GAKSPKSGAOTPKSPPEHVQETPLMFSTSVS-----SIDSPESTRSIASSVOSEPPSC 1412
QY 460 -----GQSSYPORTSMHSPPTPRRSPIPLDRPMRADSLRKOHHLDPSSAVKRTKREK 513
Db 1413 MWSGIISPOLPPSPGQTMPPSRKSTPP-----PPQTA--QTKRE- 1451
QY 514 MEMLRNDSLSQSESEVPRPPPHKSKGKMKROYSLSSSEEL-----ASTPEYSC 568
Db 1452 ---VPRKKATPAEKRES-----GPKQAVANAAYORVOVLPDADTLHFAETSPDGFSC 1502
QY 569 D-----DVEL-----ESESVESEKDSO-KGKRKT--SEOG 595
Db 1503 SSSLSALSLDEPFLQKDVLRIMPPVQENDNGNETESEQPRESENENGEKEKTIIDSKD 1552
QY 596 VLSDSN-----TRSEQRKRMVYGGHLSLEDELMSEPOIKDSCVDTCCS 639
Db 1563 LLDSDDDDDIEILEECITISAMPRTKSSRKARKPAQTASKLPPVARKPSOL-----PYK 1616
QY 640 TTLNEHSHSDKHPTVWQPSKGDRL--IGRIILN-----KLLKDGVPYRDSGAMGL 690
Db 1617 LIPSQRLQOKH-VSPTEPDDMPRYCVGEGTPINEFSTATSLSLTIESPPNELAAGBV 1675
QY 691 KYVGKMTESGRICAFITKYKKSGLADTVGHLRPGDEVLEWNGRLQGAFFEEYNIILE 750
Db 1676 R--GG--AOSG-----EFKRDITPTEG--RSTDEA-----QGGKTSVTIPELD 1714
QY 751 SKREPELVV-----SRPLGD-----IPRIDSTHAOLESSSSSESOCKMRPISVT 799
Db 1715 DNKAEGDILAEICINSMPKSGKHPRPVKIMDOVO-QASASSAPKNQOLDKKKRPT 1773
QY 800 SPKSPGLRDVPOFLSGOLSIKLMFDKVGHOLITLIGAKDLSPRDPGRPNRPVVKIYFL 859
Db 1774 SPYK-IPQNTVEYRTV-----RKMAKSKNNLNAERVF 1805
QY 860 PBRSDKNRKRTKVYKKTLEPK-----WNOTIYSPVH-----RR 893
Db 1806 SDRKSKKONLKNNSKDFNDKLPNNEDRVKSGFAFDSPHHYTPIEGTPCYFSRNDLSL 1865
QY 894 EFERRELEITLMDQAVYRESESEFLGELILELEFALLDDEPHYKIKOTHYVSLPLR- 951
Db 1866 DTDDDVDVLSR-EKALFLRKAKEKESEAVKTSHTLELTSNOOSANKTQA--IAOPINRGQ 1922
QY 952 PEPYLRROL-----HGESPTRLOR-----SKRISDEVSVDCE--- 987
Db 1923 KPIILQKOSTFPOSSKDIPDRGAATDEKLQNAIENTPVCFHSNSSLSSLDIDQENNNK 1982
QY 988 -----DCGVVSYDRHNG-----RDIQSTLSVPEOVYSSNHC 1020
Db 1983 ENBP1KETEPPOGEPSPKQASGYAPKSEFHVEDTVCFSRNSLSLSDSEDDLQEC 2042
QY 1021 SPGSPHRYVD--VIGRTSMSPSAPPPORNVBOGHRGTATGHYNTISMWD--RHRMYD 1076
Db 2043 ISSAMPKKKPSLKDNEKHS- RNM-GGLIGEDLITLDKIDQRPDEHGISPD 2095
QY 1077 HYSDDRDCCEAD-----RQ-----PYHRSRSTY 1101
Db 2096 SENFPMKAIQOEGANSIVSSLHQAAAAACLSROASSDSDSTLSLKSGLSGSPHILPPOE 2155
QY 1102 ORPL- ERTTTRSRSERDTN----- 1122
Db 2156 EEPFYSNKGPRILKGEKSTLTETKTESKGIKGGKVVYKSLITGKVSNSEISGOMQ 2215
QY 1123 -LMRSMPSLMTGRSAPPSPALSRSHPRTGSV-----OTSPTSPTGTRGRQOLPOLP 1174
Db 2216 PLQAMPPTSISRGTMIHIDGVANSSSSTSPVSKKGPLKTPASKSPSEG---QATATSP 2271
QY 1175 KOTLERSAMDIERNRQMLNKYKOVAGSDPRLEDYHSHKYSRG--WDPHRGAD----- 1226
Db 2272 RGAKPSVSELSPVARQT-----SOIGSS-----KAPSRSGSDPTPSRAQOPLSR 2319
QY 1227 TVSTSSSDSDVSAVSTSSASRFSSTSYMSVOSERRGRNKRKISVFTSKMONROMGVS 1286

```

```

Db 2320 P1QSPGRNISGPRNGISBPKNLSOLPRTSSPSTASTSYSSGSGKMS-YTS--PGROM--S 2374
QY 1287 GKNIITKSTISIGDMCSLEKNDGOSQDITAVGALGTGKKRRSSIGAKMVAIVLSKRSRA 1346
Db 2375 QONILTKQGLSKNASSIPRESASKGLNOMNNGANK-----VELSMSSYTK 2423
QY 1347 SOLQTEGGKKL-----KSTVORSTETGLAVENRNMTRQASREST----- 1388
Db 2424 SSGSESDSERPVLRQSTFIKEAPSPTLRRKLESASFSLSPPSRASPTPSQAQTPV 2483
QY 1389 -----DSMNSYS 1397
Db 2484 LSPSLPDMSLSTHSS 2498

```

```

RESULT 12
US-09-136-605-7
; Sequence 7, Application US/09136605A
; Patent No. 6140052
; GENERAL INFORMATION:
; APPLICANT: He, Tong-Chuan
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert.
; TITLE OF INVENTION: Beta Catenin, TCF-4, and APC Interact to
; TITLE OF INVENTION: Prevent Cancer
; FILE REFERENCE: 1107.75741
; CURRENT APPLICATION NUMBER: US/09/136.605A
; CURRENT FILING DATE: 1998-08-20
; EARLIER APPLICATION NUMBER: 08/821.355
; EARLIER FILING DATE: 1997-03-20
; EARLIER APPLICATION NUMBER: 09/003.687
; EARLIER FILING DATE: 1998-01-06
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 2973
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-136-605-7

```

```

Query Match 3.1%; Score 258.5; DB 4; Length 2973;
Best Local Similarity 17.9%; Pred. No. 2,9e-10;
Matches 304; Conservative 239; Mismatches 618; Indels 535; Gaps 69;

QY 30 SHLTTEERKIIILAVMDRQKKEEKGQSVLKIEHKAQPTQWPFSGITTELNNVLOPOQ 89
Db 1012 NHMDNDGELDTPINYSIKYSDQ-----LNSGRQSPQNERWARPKHILIE--DEIKQSEQ 1065
QY 90 KOPNEKEPQTKLHQEFMYKEQYVKMGEESQO--OOEQGDAPTGCIGHKTRKADCGH 146
Db 1066 ROSNQS-----TTYIVYESTDDKHLKFOPHGQOE-----CVSPKSRGANSSET 1112
QY 147 NCASYCQTKPCARCGRVSLRSNKVMVYCNLCRKQOELITKSGAMFYNSGNTLQOPOKV 206
Db 1113 N-----RVGNSNGINQNVNSQSLC-----QEDDYEDDKPTNYSERKSEERQHE-- 1155
QY 207 PGLRLEAEOEKAKIHEDPOFOGAPGDLSPVAVKGRAGHLTROPTIRNGSGVKHQIA 266
Db 1156 -----EERTPNYSIKYNEKRHRVDOPTIDYSL-----KYA 1184
QY 267 SDMPSDKRSRPSVSRDQNRREYEOSEEREDYSQYVPDQTMPRSPDYADRRSQREPOFYE 326
Db 1185 TDLPSSQKQSFSSKSSGQSSKTE-----HWSSESTSPSSNAKRONLHPSSAQ 1237
QY 327 EPGHLNVRDSNRGRHRSKE-----YIVD-----EDVE 355
Db 1238 SRGGOQKATCKVSSINQDTIQTYCEVDTPICFSRCSLSLSASADEIGCNOTQDEAD 1297
QY 356 SRQYERQREERYQARYRSDPILARYPVQPYEEDOMRIHAEVSRAHRRHSDVSLAN 415
Db 1298 SANTQIDIAIEIKETIGTRESADVPSEVPAVSOHPRTKSSRIQGS--SLSSSEARRAKAVEFSS 1356

```

| | | | | | |
|----|------|---|---------------------------------------|------------------------------|-------|
| OY | 416 | AELSDSIILLRMDRP----- | SRKSVSEBRAMNOMSVEMTREQO---- | 459 | |
| Db | 1357 | GANSFSGAQRPKSPREHYVOETPLMESCTSVS--- | SLDSFERSIASSVOSPESC | 1412 | |
| OY | 460 | -----GSSYPORTSNHSPTFRSRSPILDRPDMRADSLRKONHLDRSSAVKTRKREK | | 513 | |
| Db | 1413 | MVSGIISPSDLPDPOQOTMPPRSKIPPP----- | PPQTA-QTKRE | 1451 | |
| OY | 514 | METLNRDLSIDQSSQSVRRPPRPKHSKGMKNOVSLSSSEEL---- | ASTPEYTS | 568 | |
| Db | 1452 | -----VPKKAKPTAKRES----- | GPKOAAVNAAYORVOYLPADLTLLHFATIESTPDGSC | 15020 | |
| OY | 569 | D-----DVEL----- | ESEVSQSGSO-KGKRKT--SPQG | 595 | |
| Db | 1503 | SSSLSTSLSDPEPTQKDVLELIMPPVQENDNGNETESQPKSNNOKEAEKTIIDSKD | | 15620 | |
| OY | 596 | VLSDSN----- | TRSEROKRMUYGSHSLBEDLEMSFOIKDSQVDTSS | 639 | |
| Db | 1563 | LDDSDDDDLLEIEELISAMPTKSRKAKKQAQASLPPVAKRPOL----- | PYVK | 1616 | |
| OY | 640 | TYLNEESHSDKHPTWOPSKODRL-IGRILLN----- | KRLKDSVPRDSCAMGL | 690 | |
| Db | 1617 | LPLSONLQPKH-VSEFTPGDDMPRVYCVGEPRIINFSTATSLSDLTIESPNEIACAGV | 16750 | | |
| OY | 691 | KVNGGKTEGSRCLATITTKYKKSGLADYUHLRPGDEVLENNGRLLQCATPEEYVNIILE | 750 | | |
| Db | 1676 | R--GG-AQSG----- | BEKKDITPTG--RSTDEA----- | OGCKTSSVITPELD | 1714 |
| OY | 751 | SKPEQVELV----- | SRPIGD----- | IPRIPDTHAOLESSSSFSQKMDRSISVT | 799 |
| Db | 1715 | DNKAEBDILAEICNSAMKRGKSHKFRKTKIMDOYQ-QASASSAPKKNOLDKKKKPT | | 1773 | |
| OY | 800 | SPMSGMLRDVPQFLSGOLSIKLMFDKVGHOLIVTILGAKDLSREDRPRNPVYKIYEL | 859 | | |
| Db | 1774 | SPVKP----- | IPONTETRRV----- | RKNDSKNLNAERV | 18050 |
| OY | 860 | PDSMDKRRKTKYVKTLEPK----- | WNQFIYSPIVH----- | RR | 893 |
| Db | 1806 | SDNKDSKKOMLKNNSKDFNDKLPNNEDRYRGSPAFDPSPHNYITIGETPYCESRNDLSL | | 18655 | |
| OY | 894 | EFERMELEITLMOQARVRESESEFLIELIETLALDDEPMWYLCQNDVSSLPLPR- | | 951 | |
| Db | 1866 | DFDDDDVDLSR-EKAEUKKAKKEKESEAKVTSITELTSMQSAKNTQA-- | IAKOPINQO | 19222 | |
| OY | 952 | PSPYLPRQL----- | HGESPPRRLQR----- | SKRISDEVSDYDE-- | 987 |
| Db | 1923 | KPFLQKQSFPPQSKMDIPRGATDEKQONAIENTPYCFSHNSLSLSIDIOGENNK | | 19822 | |
| OY | 988 | -----DCGVVSDYRANG----- | RDQSSITSLVBPQVSSMHC | 10200 | |
| Db | 1983 | ENEFIKETEPDSDGERSKQASGVAPKSFHEDPYCVCSRNSLSISIEDLLOEC | | 20422 | |
| OY | 1021 | SPGSPSHRV--VIGRTSRWSPSAPPQRBVBOGHRGRATGHNTISRD-RHRYMD | | 1076 | |
| Db | 2043 | ISSAMPKPKPRSLKGNEKHSF----- | RNM-GGILGEDLTLDIKDQRPSEGLSPD | 20959 | |
| OY | 1077 | HYSSDRDRCDEAD----- | RO----- | PYHRSSTE | 1101 |
| Db | 2096 | SENDDKAIQEGANSIYVSLHQAAAACLSROASSDSILSKGISLGSFHLTPROE | | 21555 | |
| OY | 1102 | ORPLL----- | ERTTSSRSSERPDTN----- | | 11222 |
| Db | 2156 | EKPPTSXKGRILKPGEKSTLETKKIESESKIGKGGKVVYSLITGVKYSNSETIGOMQO | | 22151 | |
| OY | 1123 | LMSMSLMTGSGAPPSPLSLSRHRTQSV----- | OTPSSTPBGGRGRQRLPOLPR | 1174 | |
| Db | 2216 | PLQANMPSISRGRTMHIPEVRNSSSTSPVSKGPKLPKLTSPASKPSSEQ---- | QATATSP | 22711 | |
| OY | 1175 | KGTLESRAMDIERNOMKLNKTKYQVAGSDPRLQEDYHSHYRSQ---- | WDPIHGAD----- | 12266 | |
| Db | 2272 | RGAPSVKSELSPVYAQOT----- | SQIGGSS----- | KAPSGSGRDSPTSPKPAQPLSR | 2319 |
| OY | 1227 | TVSTKSSSDSDVSAVSTTSASRFSSTSYMSVOSERPGRNKRISVFTSKMNOBQVNS | | 12866 | |

```

Db      2320 PIGSPGRNRSISPCNGNGISPPNNKLSQLPRTTSSPTASTKSSGGGKMS-YTS-PGRQM--S 2374
QY      1287 GKNLTSTSTISGDMCSLEKRDGSDPTAVGALGTSGKRRRSIGAKMAIVGLSKRSRA 1346
Db      2375 QQNLTKOTGTSKNNASTIPRSESASKGLNOMNNGNCAKRY-----VELSMSTTK 2423
QY      1347 SOLSOTEGGCKLT-----RSTVQSTSTETLAEMNMNMTROASREST----- 1388
Db      2424 SSGSESDBSEFPVLYVROSTITKEAPEFTTLKRLKEESASELSPPSRPASPTRSQAQTPV 2483
QY      1389 -----DGSNNSTYS 1397
Db      2484 LSPSLPDMSLSTHSS 2498

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

Query Match 3 0% Score 250.5: DA.1: Length 2843;
Best Local Similarity 18.0%: Pred No. 1.1e-09;
Matches 305; Conservative 235; Mismatches 620; Indels 355; Gaps 70;

QY 30 SHLTSEERRKIIAYVDRQKEERKQSVLKIKIEHKAQPTQWPFSGITELVNNVLQPOQ 89
Db 1012 NIMDNDNGDGLDPIINVSLKYSDEQ---LNSGRQSPQSNERRARPKHIE--DEIKOSQ 1065

```


APPLICANT: CARLSON, MARY
 APPLICANT: GRODEN, JOANNA
 APPLICANT: HEDGE, PHILIP J.
 APPLICANT: JOSLYN, GEOFF
 APPLICANT: KINZLER, KENNETH
 APPLICANT: MARKHAM, ALEXANDER F.
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THILVERIS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 NUMBER OF SEQUENCES: 94
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner, Birch, Mckie & Beckett
 STREET: 1001 G Street, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20001-4598
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/452,654
 FILING DATE: 25-MAY-1995
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/741,940
 FILING DATE: 08-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A.
 REGISTRATION NUMBER: 32,141
 REFERENCE/DOCKET NUMBER: 1107,035574
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2843 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-452-654-2

Query Match 3.0%; Score 250.5; DB 1; Length 2843;
 Best Local Similarity 18.0%; Pred. No. 1,1e-09;
 Matches 305; Conservative 235; Mismatches 620; Indels 535; Gaps 70;

Db 30 SHUTEERKIIIAVMDROKKEEKEQSVLIKIEENKAQPTQWPFSGITELVNNVLQPOQ 89
 Db 1012 NIMDNDGGLDPIINSLKYSDEQ---LNSGRQSPQNERMARPKHIE--DEIKQSEQ 1065
 QY 90 KQPKNEPQTLHQOEAKTEQYKKMGESQ---QOEOKGDAPTGIGIKTKTFADGCGH 146
 Db 1066 RQSRNQS-----TTPYVYTESTDDKHLKFPFHGOE-----CVSPYKSGANSET 1112
 QY 147 NCSYQCTKFCARGGVLSRSNKKVMVNCILCRKQOELITKSGAMFYNGSNTLQOPDKV 206
 Db 1113 N-----KVGSNHGINQNVQSILC-----QEDDYEDDKPTNVSEKSEERHEE-- 1155
 QY 207 PGLRNEAPQOEKAKLHEQPOFOGAPGDLSPVAVEKGRAHGLTROPTINGSGVKNQIA 266
 Db 1156 -----EERTNYSIKYNEKRHHVDQPIDYSL-----KYA 1184
 QY 267 SDMPDRKRPSPYSRONRKYVEOSEEREDVSOYVPSDGTMPRSPDYADRRSQPOFYE 326
 Db 1185 TDIIPSSQKQSFSSKSSSGGSKTE-----HMSSSENTSTPSSNAKRONQLHPSSAQ 1237
 QY 327 EPGLNLRDNRGRHRSKE-----YIVDEDDV-----ESRDEYERQREERYQ 370
 Db 1238 SRSGQFOKATCTKVSINQETIQTYCEDVPIPCFSRCSLSLSSAEDEIGCNOTTOEAD 1297

QY 371 A-----RYRSDPNLARIPIVKQPYEQEQRTHAEVSRARHRRHSDVSLAN 415
 Db 1298 SANTLQIAETIKGIGTGRSAEDPVSEVPVAVSQHPRTKSSRLQGS--SLSESRHNAVERPS 1356
 QY 416 AELEDSRSLILMRP-----SPRQSVSEERRAENORSYMERETRENO----- 459
 Db 1357 GAKSPKSGAOTPKSPPEHYQETPLMPSRCTYS-----SLDSFSRRIASVQSEPCSG 1412
 QY 460 -----GQSSYQRTSNHSPETPRSPPIPLDRPDMRADSLRKHQHHLDPPSAVKRTKREK 513
 Db 1413 MVSGILSPSDLPDSFGCTMPSPRSKTPPP-----PPQTA--QTKRE- 1451
 QY 514 MEYMLRNDLSDDQSESVRPPPRHKKKGGKMGQVSLSSSEEL-----ASTPEYISC 568
 Db 1452 --VKNKAPTAEKRES-----GPKQAANAARVQVLPADLILFATSTPDGFGSC 1502
 QY 569 D-----DVEL-----ESESVERGDSQ--KGRKT--SEOG 595
 Db 1503 SSSLALSLDEPFIOKDYPLRLMPVQENDNGNTESEQPKESNENQEKAKTIDSEKD 1562
 QY 596 VLSDSN-----TRSEROKKMYTGGHSLLEDLNEMSEQPKINDSGVDTCS 639
 Db 1563 LLDSDDDDIETILECIIISAMPTKSSRKGGKPPAQATASKLPPVARKPSQL-----PVYK 1616
 QY 640 TILNEHSHSDKHPTWQPSKDGRL--IGRIILN-----KRLKDGSVPRDSGAMLGL 690
 Db 1617 ILPSQNRLOPQKH-VSFPTGDDMPRYCYEGTPIINSTATSLSDLTIESPPMELAGEGV 1675
 QY 691 KVVGGKMEGRICAFITVKKGSGLADIVGHLRPGDEVLEWNGRLIQGATFEVYVNIILE 750
 Db 1676 R--GG--AQSG-----EFKRDITPTGEG--RSTDEA-----QGGKTSSVTIPELD 1714
 QY 751 SKPEQVELVY-----SRIGD-----IPRIPDSTHAOLESSSSFEESQKMDPSTISVT 799
 Db 1715 DNKAEEDGILAEICINSAMKSGSHKPPRVAKKIMQYOVO-QASASSAPNNQLDGKKKPT 1773
 QY 800 SPMSQMLRDLVPOFLSGQLSIKIMLPDKVGHQILVTILGAKDLPRREDGRPNRYKATYL 859
 Db 1774 SPVKP-----IPQNTLEYRRV-----RKNAKSNMLMAERVF 1805
 QY 860 PDRSDKNRRRTKTVKTLPEK-----WNQFTIYSPVH-----RR 893
 Db 1806 SDMKDKSKQKMLKNNKSKDFNKLNNEDRVGSEFAFDPHHYTPLEGPIYCFSRNDISL 1865
 QY 894 EFEREMLEITLMDQARYESESFLGELILELEFALDLDEPHWYKIQTHVSYLPLR-- 951
 Db 1866 DFDDDVDILSR-EKAEILKAKENKESAKYTSHTELLNSQGSANKTQA--IAKQPIRNGQ 1922
 QY 952 PSPYLPRLQ-----HGESPTRIQOR-----SKRISDFVSIDYCE----- 987
 Db 1923 PKPILQOSTFPOSSKDIIPDGAATDEKLQNFALIENTPVCFSHNSLSLSDIDQENNK 1982
 QY 988 -----DGVGVSVDRHNG-----RDLSSTLSVEQVWSSNHC 1020
 Db 1983 ENEDIKETEPDPSQGEPSKDPASGYAPKSFHVEDTPTVCFSSNSSLSSLSIDSEDLLOEC 2042
 QY 1021 SPSSGPHRVD---VIGTRSWSPSAPPQORNEQCHRGTRATGHNTISMD--RARVMD 1076
 Db 2043 ISSAMPKKKRSRLKGNENKSP-----RNM--GGLGEDTLTLKDIDQPDSEGLSPD 2095
 QY 1077 HYSSDRDRDCEAD-----RQ-----PYHRSRSTE 1101
 Db 2096 SENFDWKAIQGANSIVSSLHQAANAACLRSQASSDSISLSKSGISLGSFFHLTPROE 2155
 QY 1102 QRPLL-----ERTTIRSRSESRPTIN----- 1122
 Db 2156 EKPPTYSNKGPRILKGEKSTLETKEIESESKIGKGGKVVYSLITGVYNSNEISGOMKQ 2215
 QY 1123 -LMRSMPLMTGRSAPPSPALSRSHPTGVSV-----QTSFSSITPGGRKROLPOLR 1174
 Db 2216 PLQANMPISNGRTMIHILPGVNNSSSTSPVSKGPPILKTPASKPSPEG---QATATSP 2271

```

QY 1175 KGTLEASAMDIERNRQMKLINKYQVAGSDPRLERODYHSHKYSJ---WDPHIGAD----- 1226
Db 2272 RGAKPSYKSELSPVARQT-----SQIGSS-----KAPSRSGRSDSTPSRPAQQLSR 2319
QY 1227 TVSTKSSDSDYSDYSAVSRITSSASRFSSTSYMVSQSERPRGNRKISVFTSKMQRNGVS 1286
Db 2320 PIGSPGRNSTISPGRNGISPPNKLSQLPRTSSPTASTKSSGSGKMS-YTS--PGROM--S 2374
QY 1287 GKNLTSTSTISGDMCSLEKNDGSDTAVGALGTSGKKRRSSIGAKMVAIVGLSRKRSRA 1346
Db 2375 QQNLTKOTGLSKNNASTIPRESASKGLNOMNNGANKK-----VELSRMSSTK 2423
QY 1347 SOLSOTEGGCKKL-----RSTVORSTETGLAVEMRNMTRQASREST----- 1388
Db 2424 SSGSESDRSRPPVLRQSTFIKEAPPTLRKLEBASFEESLSPSRPASPTRQAQTPV 2483
QY 1389 -----DGSNMSTYS 1397
Db 2484 LSPSLPDMSTLSTHSS 2498

```

Search completed: November 21, 2001, 16:06:17
 Job time: 353 sec

THIS PAGE BLANK (USPTO)